

# **PREDICCIÓN DE RESULTADOS DE EVENTOS DEPORTIVOS: CARRERAS DE CABALLOS**

Proyecto fin de carrera



**UNIVERSIDAD CARLOS III DE MADRID**

**ESCUELA POLITÉCNICA SUPERIOR**

Departamento de Teoría de la Señal y Comunicaciones

**Autor:** Miriam Trujillo Sevillano

**Tutor:** Fernando Pérez Cruz



Título: Predicción de resultados de eventos deportivos: carreras de caballos

Autor: Miriam Trujillo Sevillano

Director: Fernando Pérez Cruz

## EL TRIBUNAL

Presidente: \_\_\_\_\_

Vocal: \_\_\_\_\_

Secretario: \_\_\_\_\_

Realizado el acto de defensa y lectura del Proyecto Fin de Carrera el día 12 de Junio de 2012 en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid, acuerda otorgarle la CALIFICACIÓN de

VOCAL

SECRETARIO

PRESIDENTE



# Agradecimientos

Me gustaría acordarme de las personas que me han ayudado a llegar al punto en el que me encuentro actualmente.

Para empezar querría agradecer a mis padres por todo el apoyo que me han dado, por no haberme presionado en ningún momento, dándome total libertad para realizar lo que quisiera, y por ayudarme en los momentos más difíciles de la carrera.

A mis hermanos, por haber sido y seguir siendo un ejemplo a seguir para mí. Además de la ayuda que me han prestado durante muchos momentos de mi vida. Y, junto con ellos, mi abuela y mi sobrina me han dado momentos de alegría y desahogo en los momentos en los que me encontraba más agobiada.

A mi tutor, Fernando, por la ayuda y dedicación prestada durante el transcurso del Proyecto. Y a Matilde, por asumir el papel del anterior, en ausencia del mismo.

A mi novio, sin el cuál esto no habría sido lo mismo. Muchas gracias por estar ahí en todos los instantes, y darme los ánimos necesarios para confiar en mí misma.

A todas mis amigas y compañeros de clase, por todos los momentos que hemos vivido a lo largo de estos años.

Gracias a todos.



# Resumen

En la actualidad las apuestas deportivas están cobrando una gran importancia a nivel mundial. Cada vez es más común que realicen apuestas las personas menos especializadas, que ven en las mismas una vía de escape tanto como diversión como forma de hacer fortuna.

En este Proyecto Fin de Carrera nos centramos en las apuestas de carreras de caballos. Para ello vamos a tener como objetivo pronosticar qué caballo es el que va a resultar ganador en cada carrera. Y de esta forma, obtener beneficios con nuestras apuestas.

Construiremos una base de datos con carreras de Sudáfrica desde Febrero de 2011 hasta Junio de 2011, las cuáles han sido extraídas de la página web *Formstar* [1].

A través de la base de datos, entrenaremos un clasificador por medio del método de Máquinas de Vectores Soporte, con los parámetros adecuados, seleccionados con validación cruzada. De esta forma, cuando tengamos nuevas carreras, nuestro sistema será capaz de decirnos las probabilidades de que cada caballo gane la carrera, y podremos apostar con dicha información disponible.





# Índice

<b>INTRODUCCIÓN</b>	<b>1</b>
1. Objetivos	2
2. Estructura de la memoria	3
<b>CAPÍTULO 1. INTRODUCCIÓN A LAS CARRERAS DE CABALLOS</b>	<b>5</b>
1.1. Historia de las apuestas	6
1.2. Conceptos empleados	7
1.3. Categorías de carreras	10
1.4. Tipos de apuestas	14
1.5. Perfil del apostante	19
1.6. Casas de apuestas	20
1.6.1. Zeturf	21
1.6.2. Betfair	23

1.7. Artículos relacionados	23
<b>CAPÍTULO 2. MODELO</b>	<b>27</b>
2.1. Modelo elegido	28
2.1.1. Introducción	28
2.1.2. Máquina de vectores soporte (SVM) para clasificación	31
2.1.3. Pronosticando los resultados de las carreras	34
2.1.3.1. Antecedentes	34
2.1.3.2. Dos etapas previas al modelo propuesto	36
2.1.3.3. Un modelo SVM basado en dos etapas	39
2.1.4. Evaluación empírica del modelo SVM/CL	42
2.1.4.1. Datos y variables	42
2.1.4.2. Montaje experimental	43
2.1.4.3. Evaluación comparativa del modelo de dos etapas	44
2.1.4.4. Examinando el origen del beneficio	47
2.1.4.4.1. Comparación entre modelos de una y de dos etapas	47
2.1.4.4.2. Comparación de modelos lineales y no lineales	47
2.1.4.5. Debate	48
2.1.5. Conclusión	49

2.2. Modelo aplicado	50
2.2.1. Página web Formstar	50
2.2.2. Base de datos	59
2.2.3. Inferencia	69
2.2.4. Odds y Factor Q	76
2.2.5. Criterio de Kelly	82
<b>CAPÍTULO 3. RESULTADOS</b>	<b>89</b>
3.1. Elección del valor de $\beta$	90
3.1.1. Probabilidad de ganar de un caballo en función del valor de $\beta$	90
3.1.1.1. Estudio CV	90
3.1.1.2. Estudio LOO	99
3.1.2. Variación de la ganancia en función de $\beta$	108
3.1.2.1. Estudio CV	108
3.1.2.2. Estudio LOO	110
3.1.3. Elección del valor óptimo de $\beta$	112
3.1.3.1. Estudio CV	113
3.1.3.2. Estudio LOO	129
<b>CAPÍTULO 4. CONCLUSIONES Y FUTURAS LÍNEAS DE TRABAJO</b>	<b>147</b>
4.2. Conclusiones	148
4.3. Futuras líneas de trabajo	149

<b>CAPÍTULO 5. PRESUPUESTO</b>	<b>151</b>
5.1. Introducción	152
5.2. Coste del personal	152
5.3. Coste de software	153
5.4. Coste de hardware	153
5.5. Coste de material fungible	154
5.6. Resumen de costes	154

# Índice de tablas

<b>CAPÍTULO 1. INTRODUCCIÓN A LAS CARRERAS DE CABALLOS</b>	<b>5</b>
<b>Tabla 1.1.</b> Clase de carrera en función del OR	12
<b>CAPÍTULO 2. MODELO</b>	<b>27</b>
<b>Tabla 2.1.</b> Comparación entre estandarización de la carrera y estandarización de la base de datos	41
<b>Tabla 2.2.</b> Variables usadas para realizar el análisis y su definición	43
<b>Tabla 2.3.</b> Comparación empírica de los tres diferentes modelos de dos etapas a través de 156 carreras de exclusión.	44
<b>Tabla 2.4.</b> Comparación empírica de otros modelos con las 156 carreras de exclusión.	46

<b>Tabla 2.5.</b> Ejemplo matriz del caballo ganador	60
<b>Tabla 2.6.</b> Ejemplo de relleno de columnas con variable Form	61
<b>Tabla 2.7.</b> Ejemplo de relleno de columnas con las cuotas	62
<b>Tabla 2.8.</b> Correspondencia numérica con el terreno	64
<b>Tabla 2.9.</b> Resumen de los datos de la matriz numérica	65
<b>Tabla 2.10.</b> Ejemplo de matriz numérica con una carrera particular	68
<b>Tabla 2.11.</b> Ejemplo con las odds de los caballos de una carrera	78
<b>Tabla 2.12.</b> Ejemplo con las probabilidades de la página web para cada caballo en una carrera	80
<b>Tabla 2.13.</b> Ejemplo con nuestras probabilidades, calculadas a través de las odds, para cada caballo en una carrera	80
<b>Tabla 2.14.</b> Comparación de nuestras probabilidades con las de Formstar, para cada caballo, en una carrera	81
<b>Tabla 2.15.</b> Valores de las probabilidades obtenidas con SVM y de las cuotas de cada caballo en una carrera para Criterio de Kelly	84
<b>CAPÍTULO 3. RESULTADOS</b>	<b>89</b>
<b>Tabla 3.1.</b> Ejemplo de las probabilidades de ganar de cada caballo por cada $\beta$ con validación cruzada	91
<b>Tabla 3.2.</b> Ejemplo de las probabilidades de ganar para cada caballo por cada $\beta$ con Leave-One-Out	99

<b>Tabla 3.3.</b> Ejemplo de la ganancia de cada carrera en función de $\beta$ con validación cruzada	109
<b>Tabla 3.4.</b> Ganancia total para cada $\beta$ con validación cruzada	110
<b>Tabla 3.5.</b> Ejemplo de la ganancia de cada carrera en función de $\beta$ con Leave-One-Out	111
<b>Tabla 3.6.</b> Ganancia total para cada $\beta$ con Leave-One-Out	112
<b>Tabla 3.7.</b> Validación cruzada: apuestas totales y ganancias para cada tipo, con $\beta = 0$ y $\beta = 0,05$	114
<b>Tabla 3.8.</b> Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 0$	115
<b>Tabla 3.9.</b> Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 0,05$	116
<b>Tabla 3.10.</b> Validación cruzada: apuestas totales y ganancias para cada tipo, con $\beta = 0,1$ y $\beta = 0,2$	117
<b>Tabla 3.11.</b> Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 0,1$	118
<b>Tabla 3.12.</b> Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 0,2$	119
<b>Tabla 3.13.</b> Validación cruzada: apuestas totales y ganancias para cada tipo, con $\beta = 0,4$ y $\beta = 0,6$	120
<b>Tabla 3.14.</b> Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 0,4$	121

<b>Tabla 3.15.</b> Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 0,6$	122
<b>Tabla 3.16.</b> Validación cruzada: apuestas totales y ganancias para cada tipo, con $\beta = 0,7$ y $\beta = 0,8$	123
<b>Tabla 3.17.</b> Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 0,7$	124
<b>Tabla 3.18.</b> Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 0,8$	125
<b>Tabla 3.19.</b> Validación cruzada: apuestas totales y ganancias para cada tipo, con $\beta = 0,9$ y $\beta = 1$	126
<b>Tabla 3.20.</b> Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 0,9$	127
<b>Tabla 3.21.</b> Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 1$	128
<b>Tabla 3.22.</b> Leave-One-Out: apuestas totales y ganancias para cada tipo, con $\beta = 0$ y $\beta = 0,05$	130
<b>Tabla 3.23.</b> Leave-One-Out: Apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 0$	131
<b>Tabla 3.24.</b> Leave-One-Out: apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 0,05$	132
<b>Tabla 3.25.</b> Leave-One-Out: apuestas totales y ganancias para cada tipo, con $\beta = 0,1$ y $\beta = 0,2$	133
<b>Tabla 3.26.</b> Leave-One-Out: apuestas que realizamos para cada caballo, en función del tipo, con $\beta = 0,1$	134



**Tabla 3.27.** Leave-One-Out: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,2$  135

**Tabla 3.28.** Leave-One-Out: apuestas totales y ganancias para cada tipo, con  $\beta = 0,4$  y  $\beta = 0,6$  136

**Tabla 3.29.** Leave-One-Out: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,4$  137

**Tabla 3.30.** Leave-One-Out: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,6$  138

**Tabla 3.31.** Leave-One-Out: apuestas totales y ganancias para cada tipo, con  $\beta = 0,7$  y  $\beta = 0,8$  139

**Tabla 3.32.** Leave-One-Out: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,7$  140

**Tabla 3.33.** Leave-One-Out: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,8$  141

**Tabla 3.34.** Leave-One-Out: apuestas totales y ganancias para cada tipo, con  $\beta = 0,9$  y  $\beta = 1$  142

**Tabla 3.35.** Leave-One-Out: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,9$  143

**Tabla 3.36.** Leave-One-Out: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 1$  144

## **CAPÍTULO 5. PRESUPUESTO 151**

**Tabla 5.1.** Coste del personal 153

**Tabla 5.2.** Coste de software 153

**Tabla 5.3.** Coste de hardware 153

<b>Tabla 5.4.</b> Coste de material fungible	154
<b>Tabla 5.5.</b> Resumen de costes	154
<b>Tabla 5.6.</b> Presupuesto total sin y con IVA	154
<b>Tabla 5.7.</b> Plantilla del presupuesto	155

# Índice de Figuras

<b>CAPÍTULO 1. INTRODUCCIÓN A LAS CARRERAS DE CABALLOS</b>	<b>5</b>
<b>Figura 1.1.</b> Cantidad de apuestas por edades	19
<b>CAPÍTULO 2. MODELO</b>	<b>27</b>
<b>Figura 2.1.</b> Separación entre dos clases con el clasificador SVM	32
<b>Figura 2.2.</b> Ejemplo de hiperplano con función SVM	72
<b>Figura 2.3.</b> Comparación de nuestras probabilidades con las de Formstar, para cada caballo, en una carrera	81
<b>CAPÍTULO 3. RESULTADOS</b>	<b>89</b>

<b>Figura 3.1.</b> Variación de la probabilidad de ganar del caballo 1 en función de $\beta$ con validación cruzada	91
<b>Figura 3.2.</b> Variación de la probabilidad de ganar del caballo 2 en función de $\beta$ con validación cruzada	92
<b>Figura 3.3.</b> Variación de la probabilidad de ganar del caballo 3 en función de $\beta$ con validación cruzada	92
<b>Figura 3.4.</b> Variación de la probabilidad de ganar del caballo 4 en función de $\beta$ con validación cruzada	93
<b>Figura 3.5.</b> Variación de la probabilidad de ganar del caballo 5 en función de $\beta$ con validación cruzada	93
<b>Figura 3.6.</b> Variación de la probabilidad de ganar del caballo 6 en función de $\beta$ con validación cruzada	94
<b>Figura 3.7.</b> Variación de la probabilidad de ganar del caballo 7 en función de $\beta$ con validación cruzada	94
<b>Figura 3.8.</b> Variación de la probabilidad de ganar del caballo 8 en función de $\beta$ con validación cruzada	95
<b>Figura 3.9.</b> Variación de la probabilidad de ganar del caballo 9 en función de $\beta$ con validación cruzada	95
<b>Figura 3.10.</b> Variación de la probabilidad de ganar del caballo 10 en función de $\beta$ con validación cruzada	96
<b>Figura 3.11.</b> Variación de la probabilidad de ganar del caballo 11 en función de $\beta$ con validación cruzada	96

<b>Figura 3.12.</b> Variación de la probabilidad de ganar del caballo 12 en función de $\beta$ con validación cruzada	97
<b>Figura 3.13.</b> Variación de la probabilidad de ganar del caballo 13 en función de $\beta$ con validación cruzada	97
<b>Figura 3.14.</b> Variación de la probabilidad de ganar del caballo 14 en función de $\beta$ con validación cruzada	98
<b>Figura 3.15.</b> Variación de la probabilidad de ganar del caballo 15 en función de $\beta$ con validación cruzada	98
<b>Figura 3.16.</b> Variación de la probabilidad de ganar del caballo 1 en función de $\beta$ con Leave-One-Out	100
<b>Figura 3.17.</b> Variación de la probabilidad de ganar del caballo 2 en función de $\beta$ con Leave-One-Out	100
<b>Figura 3.18.</b> Variación de la probabilidad de ganar del caballo 3 en función de $\beta$ con Leave-One-Out	101
<b>Figura 3.19.</b> Variación de la probabilidad de ganar del caballo 4 en función de $\beta$ con Leave-One-Out	101
<b>Figura 3.20.</b> Variación de la probabilidad de ganar del caballo 5 en función de $\beta$ con Leave-One-Out	102
<b>Figura 3.21.</b> Variación de la probabilidad de ganar del caballo 6 en función de $\beta$ con Leave-One-Out	102
<b>Figura 3.22.</b> Variación de la probabilidad de ganar del caballo 7 en función de $\beta$ con Leave-One-Out	103

<b>Figura 3.23.</b> Variación de la probabilidad de ganar del caballo 8 en función de $\beta$ con Leave-One-Out	103
<b>Figura 3.24.</b> Variación de la probabilidad de ganar del caballo 9 en función de $\beta$ con Leave-One-Out	104
<b>Figura 3.25.</b> Variación de la probabilidad de ganar del caballo 10 en función de $\beta$ con Leave-One-Out	104
<b>Figura 3.26.</b> Variación de la probabilidad de ganar del caballo 11 en función de $\beta$ con Leave-One-Out	105
<b>Figura 3.27.</b> Variación de la probabilidad de ganar del caballo 12 en función de $\beta$ con Leave-One-Out	105
<b>Figura 3.28.</b> Variación de la probabilidad de ganar del caballo 13 en función de $\beta$ con Leave-One-Out	106
<b>Figura 3.29.</b> Variación de la probabilidad de ganar del caballo 14 en función de $\beta$ con Leave-One-Out	106
<b>Figura 3.30.</b> Variación de la probabilidad de ganar del caballo 15 en función de $\beta$ con Leave-One-Out	107
<b>Figura 3.31.</b> Variación de la ganancia total en función de $\beta$ con Validación cruzada	110
<b>Figura 3.32.</b> Variación de la ganancia total en función de $\beta$ con Leave-One-Out	112

# Introducción

Hoy en día el mundo de las apuestas abarca una gran cantidad de usuarios que ven en ellas una oportunidad de ganar dinero.

Nuestro cometido va a ser pronosticar el ganador de cada carrera de caballos para obtener un beneficio. Vamos a usar métodos de estimación de probabilidades, que nos serán útiles a la hora de calcular la posibilidad que tiene cada caballo de resultar el ganador de una determinada carrera.

Si la estimación anterior se desvía de los valores reales empleados para establecer las ganancias, existe la posibilidad de obtener un beneficio apostando sobre aquellas que son inferiores a las probabilidades estimadas.

## 1. Objetivos

El principal objetivo de este Proyecto Fin de Carrera es construir un sistema que sea capaz de, a partir de una base de datos con un determinado número de carreras, pronosticar el resultado de carreras posteriores.

Para ello, previamente tenemos que construir una base de datos que cuente con la mayoría de carreras posibles como para tener unos resultados fiables. Otra cuestión importante es que, además de gran cantidad de caballos, también interesa tener bastante información sobre las características de los mismos.

Una vez que disponemos de la base de datos, lo siguiente que tenemos que hacer es entrenar a la máquina. Para ello existen varias posibilidades, como emplear SVM (*Support Vector Machine*) o realizarlo con Procesos Gaussianos. En nuestro caso hemos elegido el primer método, ya que en él se basa el artículo que hemos seguido.

Cuando tengamos la máquina entrenada, debemos introducir las nuevas carreras (futuras) y realizar un pronóstico del resultado de las mismas.

Lo que tenemos que hacer es maximizar la ganancia, por lo tanto buscaremos los parámetros que mejor nos ayuden a ello.

El objetivo principal se puede dividir en varios más pequeños:

- Creación de la base de datos con la mayor cantidad posible de carreras pasadas.
- Estudio de distintos modelos para la posterior elección del que mejor se ajuste a nuestras necesidades.
- Implementación del modelo elegido.
- Realización de pruebas para comprobar la validez del modelo realizado.



## 2. Estructura de la memoria

La memoria está estructurada de la siguiente forma:

- Capítulo 1. Introducción a las carreras de caballos.

Haremos un recorrido a lo largo de la historia de las apuestas deportivas. Además daremos a conocer los conceptos más empleados en las carreras de caballos, así como los tipos de apuestas que se pueden dar.

Por otra parte presentaremos una serie de artículos que ayudaran a la comprensión del resto del trabajo.

- Capítulo 2. Modelo.

En este capítulo estudiaremos el modelo que vamos a seguir, y realizaremos la implementación del mismo a nuestros datos particulares.

- Capítulo 3. Resultados.

Recogeremos los resultados obtenidos aplicando el modelo definido sobre nuestra base de datos.

- Capítulo 4. Conclusiones y futuras líneas de trabajo.

En este capítulo extraeremos conclusiones a partir de los resultados obtenidos y expondremos otras líneas de trabajo que se podrían dar.

- Capítulo 5. Presupuesto.

Realizaremos el presupuesto general de los costes asociados al Proyecto. Para ello se muestran las tareas realizadas y el tiempo que hemos invertido en cada una de ellas. Además de los recursos humanos necesarios, se incluye un informe detallado de los equipos hardware.



# **CAPÍTULO 1**

## **Introducción a las carreras de caballos**

Este capítulo es fundamental para la comprensión del trabajo, ya que exponemos la temática que en él se trata.

Por otro lado, encuadramos las carreras de caballos dentro del mundo de las apuestas deportivas.

Además pondremos a disposición del lector los conceptos más importantes, que iremos empleando a lo largo del Proyecto, y una serie de estudios sobre las carreras de caballos que servirán para entender el resto del trabajo.

## **1.1. Historia de las apuestas**

Las carreras de caballos es uno de los deportes más antiguos que se conocen, y pronto la apuesta hípica estuvo ligada a él. Ya en la antigua China se apostaba por las carreras o luchas de animales, y en Arabia, la pasión por los caballos, hacía que las carreras entre ellos tuvieran un gran éxito. También en los juegos olímpicos griegos se celebraban carreras de carros, una modalidad que fue heredada por la cultura romana [2].

El auge de las carreras de caballos, y sobre todo de las apuestas relacionadas con estas carreras, tiene lugar en Inglaterra. En el siglo XX es cuando se ve un gran crecimiento de las apuestas en este país.

En Estados Unidos las apuestas llegaron debido a la influencia desde Inglaterra [3]. El origen de una apuesta deportiva se produjo cuando las carreras de caballos tenían su mayor apogeo y popularidad, especialmente en la gente de clase alta. Más tarde, con la guerra civil americana, las carreras se expandieron a todos los sectores económicos, haciendo que la popularidad aumentara de manera considerable. Sin embargo, perdieron su gran fama con la llegada de los deportes colectivos. El primero en desbancarlo fue el beisbol, seguido por el fútbol.

En España se está descubriendo en los últimos años y cada vez goza de mayor fama. Sin lugar a duda el deporte que más apuestas deportivas atrae es el deporte rey en el país, el fútbol [4]. Mucha es la gente que se atreve a realizar apuestas tanto de la liga española como de ligas extranjeras o competiciones europeas.

Hoy en día el medio preferido para realizar las apuestas en todos los países es Internet. La web nos permite realizar apuestas con un solo click, además de acercarnos las competiciones extranjeras para facilitarnos la apuesta en las mismas. Las apuestas deportivas online son sin duda un mercado con un crecimiento tremendo en nuestro país y que seguro seguirá creciendo mucho durante los próximos años [4].

## 1.2. Conceptos empleados

Para ayudar a la comprensión de este Proyecto, vamos a incluir la definición de una serie de conceptos [5][6] en orden alfabético.

**All Weather (AW).** Aquellos hipódromos que utilizan arena sintética en lugar de hierba, lo que permite las carreras de caballos pese a las inclemencias del tiempo. En el Reino Unido se usan dos tipos de compuesto: *Polytrack* (Lingfield, Wolverhampton y Great Leighs) y *Fibresand* (Southwell), produciendo este último ritmos de carreras más lentos.

**Bankroll.** Cantidad de dinero que tenemos para poder realizar nuestras apuestas.

**Blinkers.** Equipación que se coloca al caballo para restringirle la visión lateral, ayudándole a mantener la atención y evitando posibles distracciones.

**Bookie.** Persona que trabaja en una casa de apuestas fijando las cuotas en un evento deportivo. En las apuestas en directo se encarga de ir modificándolas según va avanzando el mismo. Así como de cerrar apuestas ante un hecho de relevancia para que los apostantes no apuesten conociendo el resultado (por ejemplo, que se marque un gol o que un caballo abandone la carrera).

**Bookmaker.** Casa de apuestas, es decir, lugar donde poder realizar las apuestas deportivas.

**Cuota u odds.** Cantidad que fija la casa de apuestas para cada una de las apuestas que se pueden realizar. Basando su valor en la probabilidad de que se de o no.

Ejemplo: Real Madrid (1,35) - Empate (3,15) - Numancia (5,20). Las cuotas serían cada uno de los dígitos numéricos que acompañan en cada apuesta. Existen diferentes tipos de cuotas:

- Cuota europea o decimal: Es la que se utiliza en España. Estas cuotas tienen forma de número entero o decimal (ejemplo: 1,50 o 3). Para calcular la ganancia simplemente hay que multiplicar la cantidad apostada por la cuota. Por ejemplo: si apostamos 10 euros a 1,50, el beneficio es de 5 euros.

- Cuota inglesa: Siguen el método tradicional británico. Es una cuota en forma de fracción (ejemplo: 1/2 o 2/1). Para calcular el beneficio se debe multiplicar la cantidad apostada por la cuota. Por ejemplo: 10 euros apostados a cuota 5/1, el beneficio es de  $10 \times 5/1 = 50$  €.
- Cuota americana: Siguen el método tradicional americano. Es una cuota que puede expresarse en forma de número positivo o negativo, partiendo desde el punto de vista de una apuesta de 100\$ (ejemplo: -200 o 200).

**Draw.** Sólo para carreras lisas. Se refiere al cajón de salida que ocupa el caballo antes del inicio de la carrera. El cajón de cada caballo se adjudica tras un sorteo y puede llegar a tener una importancia vital en el devenir de la carrera, dependiendo del hipódromo y la distancia a recorrer.

**Factor R.** Resistencia que tiene un caballo.

**Form.** Resultados obtenidos por un caballo en las distintas carreras.

**Going.** Término que describe el estado de la hierba del hipódromo. De más lenta a más rápida las superficies se denominan del siguiente modo: *heavy* (HY) – *soft* (SFT) - *good to soft* (GS) - *good* (G)- *good to firm* (GF) – *firm* (F). Este dato puede tener mucha importancia a la hora de seleccionar nuestra apuesta, dado que no todos los caballos se encuentran a gusto en todas las superficies.

**Handicap.** Carreras en las que el *handicapper* asigna a cada caballo el peso que debe acarrear durante la carrera, en base a los resultados cosechados por el caballo en competiciones anteriores. Se pretende que los caballos salgan del punto de salida en igualdad de condiciones. Los en teoría “mejores” ejemplares arrastran más peso mientras que los “peores” pueden correr más descargados. Cuando un caballo gana una carrera o termina al menos en un sexto puesto a partir de la tercera carrera no ganada, se le clasifica asignándole un *Official Rating* (OR). A partir del OR asignado se decide en qué nivel de carreras puede participar y qué peso debe portar en la misma.

**Handicapper.** Oficial de la *British Horseracing Authority* (BHA), encargado de decidir los *Official Ratings* (OR) y por ende los pesos que debe acarrear cada caballo en una carrera *handicap*.

**Jockey.** Se refiere al jinete que lleva el caballo.

**Official Rating (OR):** Es una forma de medir la valía de un caballo según sus actuaciones. Según el OR de un caballo podrá correr en una clase de carrera u otra.

**Racecard.** Programa de las carreras del día. En la actualidad, gracias a Internet, son interactivos y contienen información extensa sobre todas las carreras: carreras anteriores de los caballos, datos del entrenador, del *jockey*, días desde la última salida.

**Track probability.** Es la opinión de los apostantes sobre la oportunidad que tiene un caballo de resultar ganador en una carrera.

**Stake.** Es el nivel de confianza que el propio apostante concede a una determinada apuesta. El *stake* más bajo 1/10 (confianza muy pequeña o casi nula) hasta el 10/10 (nivel máximo de confianza). Dicho nivel nos marcará las pautas que deberíamos seguir con nuestro *bankroll* distribuyendo de la mejor forma el *stake* en cada una de las apuestas. Ejemplo: Si damos un *stake* 8/10 a un evento estamos otorgándole un nivel de confianza elevado y por tanto, nuestra cantidad a apostar en dicho evento también lo será.

**Starting Price (SP).** Precio de las cuotas de cada competidor justo en el momento en el que se inicia la carrera.

**Value.** Se considera este hecho cuando la casa fija una cuota desproporcionada o exagerada respecto a las probabilidades de un equipo o jugador de ganar un encuentro.

**Weight allowances.** Se refiere a determinados descuentos en el peso del caballo. Dependiendo del número de victorias del jinete, pueden pedirse rebajas.

### 1.3. Categorías de carreras

En cada hipódromo nos encontramos que en el mismo día se pueden correr varios tipos de carreras en las que participan distintos caballos, cada una de ellas tiene sus particularidades y hay que tenerlas en cuenta a la hora de apostar. En cada carrera sólo pueden participar determinados caballos que cumplan ciertas condiciones, es decir, cualquier caballo no puede realizar cualquier carrera.

En el caso de nuestra página web [1], se distinguen las carreras en [7]:

- ***Maiden***. En esta clase compiten los caballos o jinetes que todavía no han ganado ninguna carrera de las que han corrido hasta el momento. Usualmente la corren caballos de menos de 2 años, además es muy común encontrarse a algunos que aún no han realizado ninguna competición.
- ***Novice***. Agrupa a los caballos o jinetes que se encuentran en una etapa muy temprana de su carrera y que, por lo tanto, no han alcanzado todavía un determinado número de logros (normalmente no han ganado más de un par de veces).
- ***Allowance***. En este caso se produce una reducción del peso asignado a un caballo para compensarlo por la inexperiencia de su jinete.
- ***Conditions stakes***. En estas carreras el peso que debe portar el caballo depende de la edad o el sexo del ejemplar.
- ***Handicap***. Carreras en las que el *handicapper* asigna a cada caballo el peso que debe acarrear durante la carrera, en base a los resultados cosechados por el caballo en competiciones anteriores. Se pretende que los caballos salgan del punto de salida en igualdad de condiciones. Los en teoría mejores arrastran más peso mientras que los peores pueden correr más descargados. Cuando un caballo gana una carrera o cosecha al menos un sexto puesto a partir de la tercera carrera no ganada, se le clasifica asignándole un OR. A partir del OR asignado se decide en qué nivel de carreras puede participar y qué peso debe portar en la misma.



- **Stakes.** En estas carreras el dueño tiene que pagar una entrada para que el caballo pueda correr. Ese dinero se deja como bote para los que quedan en las primeras posiciones.

Hasta aquí tenemos las carreras que forman parte de nuestra base de datos, sin embargo existen otros tipos:

- Las carreras de caballos en el Reino Unido se dividen en diferentes clases. Cuanto más bajo es el número de la clase, de mejor calidad es. Así pues, los mejores caballos corren en *Class 1*, mientras que los caballos con resultados más pobres corren en clases inferiores.

La *Class 1*, se divide a su vez en tres grupos y un cuarto tipo de carreras llamado *Listed*. Así pues *Class 1*, engloba carreras de Grupo 1 (las más importantes), Grupo 2, Grupo 3 y *Listed*, enumeradas en orden de importancia.

En el caso de los saltos, la división de *Class 1*, en lugar de grupo, se llama *Grade*, así pues tratando de saltos de obstáculos la división de *Class 1* es *Grade 1* (la más importantes), *Grade 2*, *Grade 3* y *Listed*.

Estos datos son importantes de tener en cuenta especialmente cuando vamos a hacer la apuesta, pues por lo general, en las carreras de *Class 1*, no existen hándicaps, eso viene a significar que todos los caballos corren prácticamente con el mismo peso (con alguna excepción si compiten yeguas contra caballos y por la edad, ya que a los más jóvenes se les suele bonificar un poco).

Esto nos hace ver que en las carreras más importantes los caballos corren con igual de condiciones, cosa que a priori debería hacer más fácil el pronosticar el potencial de un caballo y plasmarlo en una apuesta. En cambio, en carreras de clase más baja, se suelen correr hándicaps. En este tipo de carreras, además de evaluar correctamente la destreza de un caballo, hemos de ser conscientes del peso que cargarán y que todo ello se vea reflejado en la apuesta.

En la siguiente tabla podemos ver el OR requerido para entrar a estas carreras:

Clase	OR
Class 1 - Grupo 1	-
Class 1 - Grupo 2	-
Class 1 - Grupo 3	-
Class 1 - Listed	96-110
Class 2	86-110
Class 3	76-95
Class 4	66-85
Class 5	56-75
Class 6	46-65

**Tabla 1.1.** Clase de carrera en función del OR

- Las **Nursery** son un caso concreto de hándicaps, con las mismas condiciones que las carreras de hándicap normales pero para caballos de 2 años de edad.
- Las carreras **Claimers**, son carreras por lo general de nivel bajo, donde todos los corredores están sujetos a un precio de venta establecido anteriormente a la carrera por los propietarios del caballo. Este precio ha de situarse dentro de un intervalo de precios que está determinado en las condiciones de la carrera y ligado a su vez a la calidad y al nivel de la misma. El peso que portan los caballos viene determinado también por el precio que se ha determinado para el caballo en cuestión siendo los caballos más caros los que portan más peso.
- Las carreras **Sell** suelen ser parecidas a las **Claimers**, con un funcionamiento similar salvo que el precio en lugar de fijarlo a priori, se determina por una subasta pública que se realiza después de finalizar la carrera. Son de menor nivel que las **Claimers**.
- Las carreras **Apprentice**, son carreras dedicadas exclusivamente a jinetes novatos, que aún no han obtenido 95 victorias en su carrera. Suelen ser

carreras poco frecuentes, y hay que prestar bastante atención a los *weight allowances* de los jinetes, dado que dependiendo del número de victorias que hayan cosechado en su carrera pueden pedir un descuento de peso del caballo. Este descuento se suele anotar con un número entre paréntesis al lado del nombre del *jockey*, en las *racecards*.

### Tipos de carreras según el terreno[8]

- **Flat:** Son carreras sin obstáculos o lisas. Éstas se desarrollan en hierba (*flat* convencional) o en tierra (*standard*). Las carreras *standard* no son todas iguales, ya que el material puede ser diferente y algunas pueden ser más rápidas que otras. Estas carreras también se llaman AW ya que se desarrollan todo el año. La distancia va desde 5 furlongs hasta 2 millas aproximadamente.
- **Jumps:** Son carreras con obstáculos y dependiendo del tipo de valla se dividen en dos.
  - *Hurdles*, que son vallas no muy altas y flexibles por si el caballo tropieza que no tenga que abandonar.
  - *Chase*, son las vallas altas y anchas (setos). Dependiendo del hipódromo puede ser más o menos fácil de sortear. A veces tienen agua o una altura y anchura considerable.
- **National Hunt Flat:** Es una mezcla entre carreras *flat* normal y carreras de obstáculos. Figuran en los hipódromos de vallas pero se corre sin obstáculos. Suelen ser una preparación para los caballos de cara a debutar en las vallas y su distancia es de 2 millas por regla general. Lo normal es encontrarse un favorito claro con cuota baja y muchos caballos sin datos previos.

## **1.4. Tipos de apuestas**

Como ya hemos visto, los deportes favoritos para apostar son el fútbol, el baloncesto o el tenis. Esto es así porque son los deportes más conocidos en nuestro país, de los que más información tenemos o los que nos resultan más fáciles ver día tras día. De ahí, que, por ejemplo casi todo el mundo sepa que en un partido de fútbol entre el F.C. Barcelona y el Sporting C.F. tiene muchas más posibilidades de ganar el primero.

Sin embargo, las apuestas de carreras de caballos son mucho más complejas y desconocidas, es por ello que los apostantes tienen que ser más especializados en la materia para conseguir ganar dinero.

Como normal general el orden a seguir para apostar a carreras de caballos es: decidir el dinero que se desea apostar, el tipo de apuesta y el número del o de los caballos elegidos. Si no se indica el dinero se entiende como que realizas la apuesta sobre el mínimo prefijado.

En las casas de apuestas hay tres grandes grupos de apuestas [9] [10]:

- 1. Apuesta Hípica Directa;** que incluye apostar a ganador, segundo y tercero.
- 2. Apuesta Hípica Combinada;** que puede ser apostar a exacta, imperfecta, trifecta y cuatrifecta.
- 3. Apuesta Hípica Múltiple;** donde nos encontramos con doble, triplo, cuaterna, quíntuplo, triple con canje enganche y place y cadena.

Pasamos a describir cada una de ellas:

- Apostar a ganador. Significa que se apuesta por un caballo que debe llegar primero en la carrera. Solo cobrarán si el caballo elegido gana.
- Apostar a segundo. En este caso se apuesta por un caballo, el cual puede quedar en la carrera primero o segundo. En este caso se tienen dos oportunidades de ganar.

- Apostar a tercero. Ahora las oportunidades de ganar son aún mayores, ya que el apostante cobra tanto si el caballo elegido queda primero, como segundo o tercero.

En los tres casos anteriores el dividendo se da por cada euro apostado, siendo este valor su base.

- Apostar a exacta. Hay que elegir los dos caballos que se crean que van a llegar primero y segundo en la carrera, en el orden exacto.
- Apostar a imperfecta. En este caso se escogen dos caballos que deben llegar en la carrera primero y segundo en cualquier orden.

En los dos casos anteriores el dividendo se da por cada dos euros apostados, siendo este su valor base.

- Apostar a trifecta. Se debe elegir el primer, segundo y tercer caballo en orden de llegada en el hipódromo. Se gana si los tres caballos quedan en el orden exacto que se había apostado. El dividendo se da por cada euro apostado, siendo este valor su base.
- Apostar a cuatrifecta. Consiste en elegir los caballos que finalicen primero, segundo, tercero y cuarto, en orden exacto en la carrera.
- Doble. Se seleccionan dos caballos que deben llegar primeros en dos carreras consecutivas. El dividendo se da por cada dos euros apostados en el hipódromo, siendo este valor su base.
- Triplo. En este caso son tres los caballos que se escogen. Como en el caso anterior, para que el apostante consiga el dinero, los tres caballos deben llegar primeros en tres carreras consecutivas. El dividendo se da por cada dos euros apostados en el hipódromo
- Cuaterna. Se eligen cuatro caballos para que lleguen primeros en cuatro carreras.

- Quíntuplo. Consiste en acertar el vencedor de cuatro carreras consecutivas. El dividendo se da cada diez euros apostados
- Triple con canje, enganche y place. Se trata de acertar los ganadores de tres carreras. Si gana el primero puede canjearse durante la venta del segundo pase, e igual con el tercero. De forma opcional se puede ingresar la apuesta del primero en el segundo o la del tercero, con lo que se llama “vale de enganche”. El dividendo se da por cada dos euros apostados
- Cadena. Se deben señalar seis caballos que deben llegar primeros en seis carreras del hipódromo.

Además disponemos de una serie de estrategias de apuestas que nos pueden aportar grandes beneficios [11]:

- Colocado. Apostar a colocado por un caballo supone que el mismo va a quedar entre los puestos de cabeza. El número de puestos a tener en cuenta depende del número de participantes que tenga la carrera. Usualmente en carreras de 4 corredores o menos esta opción no está disponible. En carreras de 5 a 7 corredores, el colocado son el primer y segundo caballos en llegar a meta, mientras que en carreras de 9 a 15 son los tres primeros y en carreras de 16 o más corredores suelen ser los 4 primeros. Esto no es siempre cierto y es algo que hemos de leer en las reglas de cada carrera, puesto que en carreras tipo *Maiden* independientemente del número de caballos, solo se tiene como colocado si queda entre las tres primeras posiciones.
- *Back* o *Lay*. Este tipo de apuestas se refiere a apostar a favor (*back*) o en contra (*lay*) de un caballo. Apostar a *backs* suele ser más complicado puesto que hay que acertar el ganador o colocado. Si se opta por los *lays*, con que nuestro caballo no gane ya tendremos la apuesta acertada. Sin embargo, apostar a favor reporta mayores beneficios, ya que a mayor riesgo mayores ganancias.

- *Lay placed*. Estrategia de apostar en contra de que el caballo favorito termina en los puestos de colocado. En las carreras de caballos es muy frecuente encontrar sorpresas, ya que hay muchos factores que pueden influir en el desarrollo de la carrera como el estado del suelo, la posibilidad de que sufra un tropezón o caída e incluso el estado en el que se encuentre el caballo en ese momento. De ahí que sea muy rentable el apostar en contra de esos caballos que la mayoría tiene como favoritos, ya que podemos aspirar a un gran premio apostando una cantidad pequeña.

Tener en cuenta que en carreras de vallas, que suelen ser más accidentadas, o en carreras largas, esta estrategia puede llevar a mayores beneficios.

- Apostar por el jinete. Últimamente se está haciendo común pensar que tiene más ventaja apostar por el jinete corredor que hacerlo por el caballo. Además es mucho más sencillo hacerse con los datos de los jinetes que interpretar el pedigree, la crianza o el terreno para un solo caballo. De ahí que para los apostantes que no tienen mucho conocimiento sea una buena estrategia de apuestas.
- Las apuestas al totalizador o tote [12]. Es un bote donde las cuotas no se conocen nunca hasta el momento en el que se cierran las apuestas. Así pues, nuestra cuota vendrá determinada por el volumen de dinero que ha entrado a cada uno de los participantes de la carrera. Cuanto más dinero entre por nuestro caballo, menor cuota nos pagarán, pues es inversamente proporcional. Además del bote total, cabe descontar la comisión que se queda el tote, como organismo para sufragar gastos y generar su beneficio, pues a fin de cuentas es otra forma de negocio. Apostar al tote, suele ser interesante cuando tenemos la sospecha de que nuestro caballo no va a ser muy popular, ya que su cuota será alta, y es probable que en el tote se acabe pagando una cuota mayor. Esto es válido tanto para las apuestas de ganador como para las de colocado. Sin embargo, con la aparición de las casas de intercambio de apuestas, estas ventajas han ido reduciéndose mucho, llegado el punto que no compensa demasiado apostar en el tote, al menos a lo que se refiere a cuotas de ganador y colocado. La apuesta mínima en este tipo de

combinación es de dos euros o dos libras, dependiendo del país en el que estemos realizando la apuesta. Obviamente, como está previsto que haya muchos acertantes, el pago que recibiremos como compensación es bastante reducido.

- Estrategia del *trading*. Para comprenderla mejor vamos a poner el ejemplo de la compra-venta inmobiliaria. La operación perfecta para obtener beneficios es comprar una casa con el propósito de venderla posteriormente a un precio bastante más elevado. En las apuestas de caballos se denomina *trading* a la obtención de un beneficio derivado de apostar a favor de que un caballo ganara una determinada carrera (*back*), para luego, a lo largo de la carrera apostar en contra de ese mismo caballo (*lay*). De esta forma nos aseguramos que obtendremos beneficios pase lo que pase en la carrera.

Estas estrategias pueden ser llevadas a cabo antes del inicio de la carrera o durante la misma (*in-play*).

Otra cosa a tener en cuenta es que en las carreras de caballos suelen ser muy frecuentes los retiros [13]. El motivo principal suele ser que no entre en el cajón de salida (en el caso de antes de la carrera) o que ocurra algo durante el transcurso de la misma que lo haga abandonar. Las retiradas minimizan las cuotas. De esta forma, si el caballo retirado es por el que habíamos apostado se nos devuelve el dinero jugado, si es otro caballo se produce una reducción de nuestra cuota.

En algunas casas de intercambio cuentan con un factor de reducción para los caballos (en tanto por ciento), al cual tiene acceso el apostante. Sin embargo, en las casas tradicionales la reducción se hace en función de la cuota que tenga el caballo en el momento del abandono.

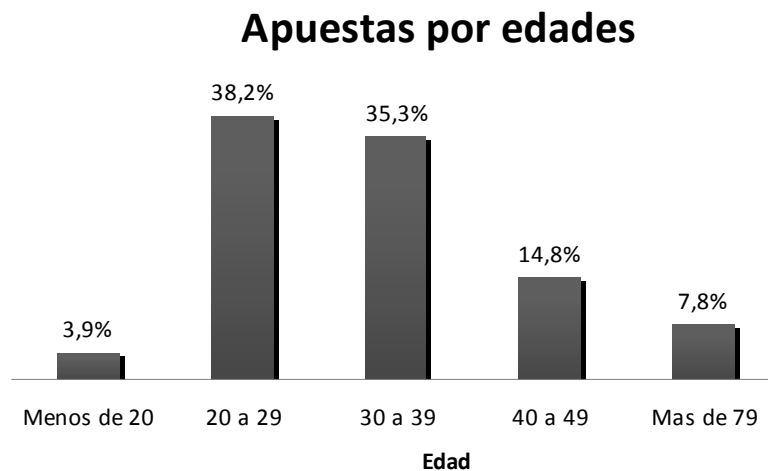


## 1.5. Perfil del apostante

Según una encuesta realizada en el País Vasco [14] por las tres casas de apuestas que operan en la CAV (Kiroljokoa, Reta y Garaipen-Victoria), el perfil del apostante, a grandes rasgos, es el de un varón de entre 25 y 45 años, con gran afición por el deporte, lector habitual de prensa especializada y, por lo tanto, conocedor de la actualidad deportiva. Además, suelen ser usuarios cotidianos de Internet, y muchas veces realizan las apuestas en compañía de amigos.

Un estudio realizado por *Betfair* [15] nos dice que de los apostantes, las mujeres apenas superan el 15%.

En la siguiente gráfica de *Paysafecard* [16] (compañía líder en Europa en la comercialización de cupones prepago para compras online y miembro de la AEDAPI) podemos ver el reparto por edades:



**Figura 1.1.** Cantidad de apuestas por edades

En cuanto a la cantidad apostada, normalmente no superan los 10€ cada vez que juegan. Además, el 60% juega entre 2 y 4 veces por semana, mientras un 30% lo hace sólo 4 veces al mes.

En España, el deporte estrella es el fútbol en todas sus variantes, donde se apuesta desde quién marcará el primer gol hasta cuántos minutos descontará el arbitro, seguido por el tenis, por el cual se tiene preferencia a realizar las apuestas siguiendo la retransmisión en directo.

Tenemos que irnos a Inglaterra o Australia para encontrarnos con que las carreras de caballos o el rugby superan el 50% de las apuestas.

### **Postura a elegir cuando apostamos [17].**

A la hora de apostar podemos elegir entre dos perfiles básicos

- Apostante conservador. En este punto nos encontramos con los apostantes que no realizan apuestas superiores a una cuota de 1,5, de ahí que las ganancias sean más regulares, pero que los beneficios sean bastante pequeños. Además sus apuestas siempre estarán acordes con toda la información de que disponen.
- Apostante arriesgado. En este caso el apostante huye de las cuotas inferiores a 2,00. Perderá en más ocasiones que el anterior, pero cuando gane lo hará con una cantidad mucho mayor. Este tipo de apostante se deja llevar más por impulsos o por el corazón.

## **1.6. Casas de apuestas**

El concepto de casas de apuestas puede abarcar dos grandes grupos, los *bookies* tradicionales y las denominadas casas de intercambio de apuestas [18] [19].

Dentro del grupo de *bookies* tradicionales se encuadran la mayoría de las casas de apuestas que actualmente existen en el mercado. En ellos, la forma de negocio es interactuar directamente con el apostante, haciéndose cargo del pago de la apuesta si esta es ganadora o quedándose con el importe del apostante si esta resulta perdedora. Son grandes grupos inversores los que se encuentran detrás de estas, creando de ellas importantes multinacionales. El modo de encontrar beneficios para este tipo de casas resulta de la aplicación de lo que llamamos cuotas. Esto es, se trata de jugar con las cuotas, de manera que los ingresos por apuestas perdedoras de los apostantes de un resultado nunca sean inferiores a los pagos de apuestas ganadoras de los apostantes del resultado correcto. Esto se hace no repartiendo nunca el 100% de lo recaudado en lo apostado. Así, pase lo que pase en un juego, la empresa dispondrá de beneficios. En este

grupo se integran muchas de las grandes casas de apuestas conocidas: Bwin, bet365, Miapuesta, Interapuestas, William Hill, etc.

En el caso de intercambio de apuestas, el *bookie* actúa de intermediario entre 2 apostantes. Ellos serán los que se pondrán de acuerdo tanto en la cantidad a apostar como en las cuotas del evento en cuestión. Aquí, la casa de apuestas, obtendrá el beneficio a través de una comisión de un porcentaje que le será cobrada al usuario que termine siendo el ganador de dicha apuesta. Este porcentaje oscila en torno a un 3% y un 5% dependiendo de la casa de apuestas. En este grupo podemos encontrar como principal exponente *Betfair* y otras menos potentes como *RedBet* o *Betsson* (esta última también ofrece servicios como una casa tradicional).

Normalmente suele ser más rentable para el apostante el tipo de intercambio, aunque también suele ser un poco más complejo de utilizar y requiere algo más de experiencia.

En el mundo de las apuestas, la reputación e integridad de un corredor de apuestas es su principal valor y esto hará que los apostantes sigan confiando en esa casa de apuestas o no, para futuros juegos [20].

Con la llegada de Internet muchos corredores de apuestas tienen una página en línea donde se pueden ver eventos deportivos y apostar en vivo. De estas páginas vamos a destacar las siguientes:

#### **1.6.1. ZEturf**

Fundada en Francia en junio de 2005 [21]. Fue la primera casa de apuestas hípcas on line. Es considerada como la primera oferta completa que abarca todo el mercado europeo de carreras hípcas internacionales consiguiendo un promedio de más de trescientas mil visitas diarias.

Zeturf sólo es un mediador entre el apostante y la apuesta garantizando el pago y transparencia en la misma.

Para dotar de un mayor grado de seguridad, tiene unas reglas estrictas de inscripción, inflexibilidad en las políticas de depósitos y retiradas, todo supervisado por

*Barclays Bank* que se ocupa de las transacciones financieras. Además entre sus socios cuenta con *Gamcare*, que es un organismo internacional que se encarga de supervisar lo relacionado con los abusos en juego de dinero, asegurando de esta forma su transparencia.

Además en la página web [22] contamos con toda la información de los caballos, acceso a los pronósticos, vídeos de las carreras y resultados.

Por lo tanto, podemos decir que esta casa de apuestas nos da mucha seguridad y tiene el respaldo financiero adecuado a la par que ofrece multitud de datos para ayudarnos a realizar las apuestas.

En ZEturf tienen sus propias apuestas disponibles [23]:

- La carrera principal del día: el ZE5 (quinté+). En el ZE5, debes encontrar los 5 primeros caballos de la carrera. La carrera se desarrolla cada día a las 13h50 hora francesa de lunes a viernes y a las 15h08 los sábados y domingos.
- El ganador (apuesta dicha S). Esta apuesta cubre el *Gagnant* y el *Placé*. Si apuestas sobre un caballo *Gagnant*, este caballo tiene que ganar la carrera. Si apuestas *Placé*, tiene que acabar la carrera entre los tres primeros puestos o primero o segundo si hay menos de ocho caballos.
- La gemela (apuesta dicha J). Esta también puede ser *Gagnant* o *Placé*. Aquí eliges dos caballos. Si apuestas *Gagnant*, tus dos caballos deben de acabar primero y segundo. Si apuestas *Placé*, tus dos caballos deben de acabar la carrera en los tres primeros lugares.
- La exacta (apuesta dicha Ju). Aquí eliges dos caballos, y deben de acabar la carrera en los dos primeros lugares, en el orden elegido.
- El trio (apuesta dicha Tr). Eliges tres caballos, y estos deben de ser los tres primeros de la carrera.

- El cuarto (apuesta dicha Zc). Aquí debes de encontrar el caballo que va a acabar cuarto de la carrera.
- La trífecta (apuesta dicha Tri). Se eligen tres caballos, y deben de acabar la carrera en los tres primeros lugares, en el orden que dijiste.
- El ZE 2/4 (apuesta dicha ZE 2/4). Eliges dos caballos, y tus dos caballos deben de acabar la carrera en los cuatro primeros lugares.
- El cuarteto (apuesta dicha ZE4). Debes de encontrar los cuatro primeros caballos de la carrera.

### 1.6.2. Betfair

Fue fundada en 1999, siendo hoy en día una de las casas de apuestas más importante en el mundo [24]. Es un gran ejemplo de intermediario entre jugadores, ya que en *Betfair* las cuotas son fijadas por los usuarios, al apostar unos contra otros, y no por la casa misma.

El único cometido de *Betfair* es encargarse de que las operaciones se lleven a cabo, por ello se queda con un 5% de las ganancias de los apostantes.

Para las carreras de caballos, uno de los métodos más comunmente utilizados, es realizar el *trading*, que, como hemos explicado los tipos de apuestas, consiste en apostar a favor de la victoria de un caballo para posteriormente apostar en su contra.

Además en su página web [25] podemos encontrar información relativa a casi todos los deportes existentes, así como la posibilidad de realizar apuestas en los mismos.

## 1.7. Artículos relacionados

En este apartado vamos a describir una serie de artículos que nos ayudarán a entender mejor el modelo en el que nos vamos a basar.

Existe un artículo [26] que investiga las estrategias fundamentales para detectar y explotar los errores de los sistemas públicos en los mercados de apuestas de carreras de caballos.

La realización del modelo de *Multinomial Logit* para los procesos de carreras de caballos se estima a partir de una base de datos de 200 carreras. Para obtener una estimación más eficiente de los parámetros se emplea un procedimiento recientemente desarrollado que explota el contenido de la información en el conjunto de órdenes de finalización.

Entre las variables de este modelo de probabilidad se incluyen las características del caballo y del jinete, además de varias características específicas de la carrera. Además se emplean procedimientos de muestreo y retención para evaluar las estrategias de apuestas.

Los estudios para evaluar la eficiencia de los mercados de apuestas de carreras de caballos están motivados por las similitudes entre estas y los mercados de valores, tales como incertidumbre en los beneficios futuros de las inversiones, presencia de muchos participantes, o la disponibilidad de la variedad de información relativa a las investigaciones y a los participantes [27][28][29].

Este artículo busca un sistema de apuestas rentable aplicando la configuración de apuestas mutuas. Los sistemas de apuestas tienen dos componentes: un modelo del proceso de carreras de caballos, y una estrategia de apuestas. El modelo de procesos de carreras se basa en predecir el resultado de la carrera. Su base es predecir la probabilidad de cada caballo de ganar la carrera. La estrategia de apuestas utiliza esas probabilidades como entrada a un algoritmo de apuestas el cual determina las cantidades a apostar por cada caballo.

En otro artículo [30] se presenta el resultado de dos años de aplicación del método de validación cruzada de un sistema de hándicap.

Se basa en el rendimiento de los datos de las carreras de toda una temporada siendo sometidos a un análisis discriminante y un criterio de clasificación. La función

discriminante y el criterio de clasificación del primer año (Año 1) se aplica a los datos del segundo año (Año 2). Las técnicas de clasificación son evaluadas en términos de porcentaje de clasificación correcto y de retornos de la inversión.

Este estudio está diseñado para medir la eficacia relativa del método de validación cruzada en seis distintas técnicas de clasificación. El término de discriminación se refiere al proceso estadístico que aprovecha los sistemas de carreras hándicap que es capaz de diferenciar entre ganadores y no ganadores. El término de clasificación se refiere a la aplicación de las reglas iniciales a un segundo ejemplo de carreras.

Los datos se obtienen a través de la observación del rendimiento de caballos ganadores y perdedores en dos temporadas (referidas como Año 1 y Año 2). No está asumiendo que los dos encuentros dan unos resultados únicos que no podrían haber ocurrido durante otra temporada. El problema de discriminación se basa en la existencia de una función lineal que produce una separación significativa entre ganadores y no ganadores basados en los datos del Año 1. El problema de clasificación toma la función discriminante y el criterio de clasificación el Año 1 y realiza la validación cruzada con los datos del Año 2.

Los datos se obtienen de dos temporadas del *Golden Bear Raceway* en Sacramento. Carreras realizadas entre Mayo y Julio. Todas son de una milla de distancia y con ocho o nueve caballos por carrera

Los estudios de la eficiencia de los mercados de apuestas de carreras normalmente examinan los retornos de las apuestas ganadoras [31]. La conclusión de esta tarea es que existen tendencias en las probabilidades de ganar, además, esas tendencias son insuficientes para crear una estrategia rentable, excepto en raras ocasiones donde hay un claro favorito.

Para examinar los retornos de las apuestas, se requiere estimar la probabilidad de cada resultado en las carreras. Hausch, Ziemba, y Rubinstein (1981) aplicaron el modelo descrito por Harville (1973) para obtener tales estimaciones. Las fórmulas de Harville tienen limitaciones descritas por Hausch et al. En este artículo se describen una serie de alternativas al modelo de Harville. Las fórmulas resultantes derivadas de los

modelos alternativos renuncian a la simplicidad, del modelo de Harville, a cambio de estimaciones más precisas.

En este artículo también se describen los mercados de apuestas y los estudios realizados sobre la eficiencia de los mercados. En otras secciones describen las fórmulas de Harville, las fórmulas alternativas y la comparación con los dos modelos.

Las carreras de caballos tienen grandes similitudes con la inversión en el mercado de valores [32]. En ambas situaciones las ganancias futuras no son conocidas con certeza, hay un gran número de participantes, mucha información disponible, abunda el asesoramiento profesional, y cada participante tiene información sobre las actividades de los otros apostantes (o inversores). Los datos de las carreras de caballos nos permiten tomar ciertas actitudes con respecto al riesgo y el comportamiento de inversión.

Este artículo estudia un nuevo ejemplo de resultados de carreras de caballos de Atlantic City, New Jersey. Las cuestiones que estudian son:

1. ¿Son las apuestas de mercado, determinadas por el comportamiento de apuestas del público, los mejores datos para realizar la predicción del orden de finalización?
2. ¿Las apuestas de salida de cada caballo refleja la verdadera probabilidad de ganar de ese caballo como sugiere Baumol (1965) o hay una tendencia sistemática por apostar en exceso por los caballos favoritos como sugieren Rosett (1977), Ali (1977) y Snyder (1978)?
3. Ya que cada apuesta se registra casi inmediatamente después de ser realizada, puede haber ventajas para las personas que realizan las apuestas minutos antes de que comience la carrera. Esta estrategia minimiza el tiempo en que la señal producida por la apuesta estará disponible para otros apostantes. ¿Hay alguna evidencia de que los apostantes finales tienen mejor información que los iniciales?



# **CAPÍTULO 2**

## **Modelo**

A pesar de que las carreras de caballos no son uno de los deportes preferidos para realizar apuestas, existen numerosos modelos que nos pueden ayudar a realizar una mejor estimación de la probabilidad que tiene un caballo de ganar.

A continuación, vamos a describir el modelo que nos ha ayudado a llevar a cabo nuestro Proyecto.

Una vez descrito el modelo elegido, pasaremos a implementarlo en nuestro ejemplo. De forma que seguiremos las pautas marcadas por el mismo, pero con el ejemplo que nos atañe.

## 2.1. Modelo elegido

### **Identificando a los ganadores de eventos competitivos: un modelo de clasificación basado en SVM para la predicción de carreras de caballos [33]**

El objetivo del modelado de las carreras de caballos es evaluar la eficiencia de la información que hay disponible en los mercados de apuestas.

Lo habitual es hacer un pronóstico de las posiciones finales de los corredores a través de modelos de regresión discretos o continuos. Sin embargo, hay evidencias empíricas que indican que la información contenida en las posiciones finales suele ser poco fiable.

Para resolver este problema, se propone una clasificación basada en el modelado de paradigma que se fundamenta solo en datos que distinguen a los ganadores de los perdedores. Para evaluar la eficacia de estos modelos se realizó un experimento usando datos de carreras de Inglaterra.

#### 2.1.1. Introducción

Muchos estudios exploran los mercados de apuestas de carreras de caballos porque comparten muchas características en común con mercados financieros más amplios, incluyendo un gran número de participantes y una amplia gama de factores que pueden influir en los caballos [34][35][36][37][38][39][40]. Además, los mercados de apuestas ofrecen una importante ventaja sobre los mercados financieros más amplios, esto es, generan un resultado inequívoco (un ganador) y una tasa de retorno asociada dentro de un marco de tiempo finito [36], y por lo tanto proporcionan un punto de referencia objetivo para medir la calidad de decisión de inversión (ejemplo una apuesta). Como resultado, los mercados de apuestas pueden proporcionar una visión clara de las cuestiones de fijación de los precios que son más complicados en otros lugares [38] y el valor de estudiar las decisiones del apostante es reforzado por el hecho de que estos mercados son en sí mismos importantes. Por ejemplo, el volumen de negocio del

mercado de apuestas de caballos en Reino Unido en 2006 fue de 19.200 millones de euros.

El modelo de predicción es empleado cuando hay que calcular el grado con el que los apostantes hacen un uso eficiente de la información al hacer sus decisiones de inversión. Particularmente, los modelos incorporan variables basadas en la información pública disponible, empleada para estimar las posibilidades de los caballos de ganar. Si estas estimaciones permiten hacer apuestas rentables sobre un gran número de carreras futuras, podemos concluir que los apostantes no descartan totalmente la información relativa a los atributos contenidos en el modelo [26][35][41][42].

Se ha demostrado que pronosticar el ganador de una carrera es importante para explicar la fuerza relativa de los competidores dentro de la carrera [26]. El modelo *Conditional Logit* (CL) [43] se ha propuesto para esta tarea, ya que, a diferencia de la regresión logística común que considera a cada caballo por separado en cada carrera, CL modela una carrera como una entidad y, por consiguiente, mantiene una relación entre los corredores que compiten [26][44][45]. Recientemente, Edelman (2007) [46] demostró que la exactitud de predicción de estos modelos se puede mejorar si se usan junto con métodos más modernos. Este enfoque está basado en la filosofía de dos etapas de Benter (1994) [41] y utiliza la Regresión de Vectores Soporte (SVR) para modelar la relación entre:

- Variables fundamentales asociadas con las actuaciones recientes de los caballos y factores relativos a las carreras actuales (ej: premio en dinero, peso transportado...).
- Posición final de los caballos.

Las previsiones resultantes se combinan con los precios finales de los caballos del mercado (*odds*) mediante CL en un segundo paso. CL y SVR se complementan entre sí en el sentido de que CL representa la competencia dentro de la carrera, mientras que la SVR se adapta a una gran cantidad de variables potencialmente correlacionadas, con bajo riesgo de sobreajuste y modelos complejos con relaciones no lineales entre atributos, de manera controlada por datos.

Este artículo desarrolla un modelo de pronósticos que adopta las dos etapas del modelo de aproximación. Sin embargo, mientras trabajos previos en pronósticos de carreras de caballos se basan sobre todo en métodos regresivos [46][47], el modelo aquí propuesto expresa Máquinas de Vectores Soporte (SVM) para clasificar los resultados de las carreras. Como se explicará más adelante, consideraciones teóricas, así como los resultados empíricos [48] arrojan dudas sobre la fiabilidad de un componente clave del modelo de regresión, el orden de importancia de los datos del final. Tomando una aproximación de clasificación, el modelo se centra en distinguir caballos ganadores y no ganadores y evitar el uso excesivo de rankings incorrectos, especialmente en las posiciones más bajas de la clasificación. Además, un nuevo modelo de pre-procesamiento de datos es aconsejado para introducir algunas ideas de la competencia dentro de la carrera, en la fase del modelado de la primera etapa.

El objetivo principal de este artículo es examinar la eficacia del método de clasificación propuesto para la predicción de carreras de caballos, y arrojar luz a la contribución marginal de los elementos de este modelo de dos etapas. Para terminar, una evaluación empírica es necesaria para contrastar el rendimiento predictivo propuesto por el modelo basado en SVM con puntos de referencia muy competitivos [42][46]. Los componentes del modelo predictivo (la aproximación jerárquica de dos etapas, los nuevos datos de la técnica de pre-procesado y el modelo no lineal) son evaluados de forma individual para confirmar su adecuación.

El resto del artículo se organiza así: la teoría del SVM revisada antes, describiendo el modelo de las carreras de caballos y los detalles de las dos etapas del modelo. Por último resultados de la evaluación empírica y conclusiones.

### 2.1.2. Máquina de vectores soporte (SVM) para clasificación

El SVM es una técnica de aprendizaje que facilita la clasificación binaria lineal y no lineal.

Por ejemplo:

$$S = \{(x_i, y_i)\}_{i=1}^M$$

Donde:

- $x_i \in X \subseteq R^N$  siendo  $x$  un vector de  $N$  dimensiones
- $y_i \in \{-1, +1\}$  la correspondiente etiqueta binaria
- $M$  es el número de observaciones

SVM aprende a partir de los datos de un modelo funcional.  $f_\Lambda(x) : X \mapsto \{-1, +1\}$ . Esto permite la estimación de la pertenencia de las observaciones no contenidas en  $S$ .

- El vector  $\Lambda$  incluye los parámetros del clasificador que son adecuados en  $S$  en una etapa de la construcción de los modelos (clasificador de entrenamiento).

La SVM está inspirada en la teoría del aprendizaje estadístico [49]. Para llegar a un modelo de clasificación de  $S$ , se implementa el concepto de máximo margen de separación. Esto es, se intenta maximizar la distancia entre casos que se encuentran cercanos a la línea de decisión que separa las dos clases [50]. Puede verse que maximizando este margen el error en los datos futuros se minimiza [49].

Para construir este clasificador lineal con máximo margen, la norma correspondiente al hiperplano del vector de pesos,  $w$ , debe ser minimizado. Sujeto a la restricción de que los datos de entrenamiento de cada clase residen en cada lado de la superficie de separación.

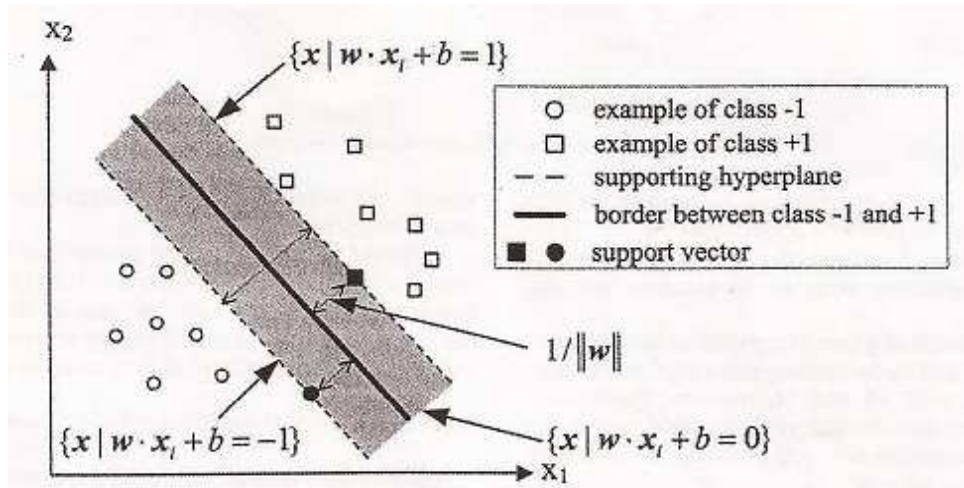


Figura 2.1. Separación entre dos clases con el clasificador SVM [33]

Con  $y_i \in \{-1, +1\}$ , esta restricción puede ponerse como [51]:

$$y_i((w \cdot x_i) + b) \geq 1, \quad i = 1, \dots, M \quad (2.1)$$

Los ejemplos que satisfagan la ecuación (2.1) se llaman vectores soporte y definen la orientación del hiperplano.

Para evaluar los errores de clasificación (esto es, ejemplos que no cumplen (2.1)), se introducen variables auxiliares que relajan la condición en (2.1) [50],  $\xi_i$ . Por lo tanto, para construir un clasificador SVM de máximo margen, tenemos que resolver el siguiente problema de programación cuadrática convexa:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^M \xi_i \\ \text{s.t.} \quad & y_i((w \cdot x_i) + b) \geq 1 - \xi_i, \quad i = 1, \dots, M \end{aligned} \quad (2.2)$$

Las variables  $w$  y  $b$  definen la separación del hiperplano, de modo que el clasificador resultante tiene la forma:

$$f_{\Lambda=\{w, b\}}(x) = \text{sign}((w^* \cdot x) + b^*) \quad (2.3)$$

Donde  $w^*$  y  $b^*$  representan la solución de la ecuación (2.2).

Para construir una superficie de decisión general no lineal, el mapa de SVM tiene los datos de entrada en un espacio de características de alta dimensión a través de una función de características de asignación,  $\Phi$ . La construcción de un hiperplano de separación en este espacio de características lleva a un límite de decisión no lineal en el espacio de entrada [49]. Se ha demostrado, con conjuntos de datos conocidos, la capacidad de la SVM para revelar relaciones no lineales entre variables de entrada mediante la proyección de los datos en un espacio de dimensión superior [52]. Por ejemplo, la clasificación no lineal estandar, como el problema XOR o la clasificación de un tablero de ajedrez en las regiones en blanco y negro se resuelven con una SVM [50][53][54].

El mapeo de los datos se lleva a cabo para evitar muchos calculos en el espacio de características transformado. Considerando la ecuación (2.2) con  $\alpha_i$  denotando los multiplicadores de Lagrange [49][51]:

$$\begin{aligned} \max_{\alpha} &= \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j (x_i \cdot x_j), \\ \text{s.t.} : & \sum_{i=1}^M \alpha_i y_i = 0; \quad 0 \leq \alpha_i \leq C \quad \forall i = 1, \dots, M \end{aligned} \quad (2.4)$$

La ecuación (2.4) contiene los datos de entrada en forma de productos escalares. Se le puede llamar función *Kernel*,  $K$ , y puede ser empleada para calcular el producto escalar de los vectores transformados en el espacio de entrada:

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (2.5)$$

El *Kernel* puede considerarse como una función para medir la distancia entre dos vectores de entrada en el espacio no lineal. El clasificador resultante es:

$$f_{\Lambda=\{\alpha,b\}}(x) = \text{sign} \left( \left( \sum_{i \in SV} \alpha_i y_i K(x_i, x) \right) + b \right), \quad (2.6)$$

$$\text{Con } SV = \{i/\alpha_i > 0\}$$

Donde el conjunto SV contiene los vectores soporte.

En este artículo, se emplea la función *Kernel* Gaussiana de base radial (RBF)

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \quad (2.7)$$

Que además presenta menos dificultades numéricas, ya que los valores de salida de las funciones Gaussianas se encuentran entre 0 e  $\infty$  [49].

Con el fin de resolver una tarea de clasificación con RBF-SVM, hay que determinar dos parámetros:

- El parámetro de regularización,  $C$ , que controla el equilibrio entre la maximización del margen y la clasificación del conjunto de entrenamiento sin error.
- El parámetro de suavizado,  $\gamma$ , que determina el ancho de la Gaussiana, y con ello la sensibilidad de la medición de la distancia.

Estos parámetros se calculan a través de la selección de distintos valores para  $C$  y  $\gamma$  evaluando de forma empírica el resultado con todas las combinaciones posibles.

### 2.1.3. Pronosticando los resultados de las carreras

#### 2.1.3.1. Antecedentes

Los modelos predictivos ayudan a examinar la eficiencia de los mercados de apuestas de carreras de caballos. La opinión de los apostantes sobre la oportunidad de un caballo de ganar se llama *track probability*,  $q_i^j$ , y puede obtenerse por las probabilidades de ganar del caballo  $i$  en la carrera  $j$ ,  $u_i^j$ , por medio de  $q_i^j = \frac{1}{1 + u_i^j}$ . Un

mercado es eficiente, desde el punto de vista de la información, si los apostantes tienen en cuenta toda la información que hay disponible en el mismo. Las *odds* representan la mejor estimación que hace el mercado de la posibilidad de un caballo de ganar, y deberían reflejar la verdadera probabilidad de ese caballo en esa carrera. Las apuestas sólo serán rentables si las probabilidades dadas por las casas de apuestas son inexactas. El objetivo de ese modelo es evaluar con precisión estas probabilidades basadas en la



información pública disponible. Si se demuestra que las apuestas sobre las bases de estas probabilidades obtienen un beneficio, se puede decir que el modelo tuvo éxito al poner la información a disposición de los apostantes.

La toma de decisiones al apostar en las carreras de caballos puede ser modelada como un proceso discreto. El modelo de CL [43] ha aparecido como una aproximación para estudiar las preferencias de los apostantes entre los diferentes participantes de la carrera. A diferencia de las regresiones lógicas ordinarias, las cuales tratan cada dato (cada caballo) individualmente, CL mantiene las conexiones dentro de las alternativas de un conjunto de elección (esto es, entre los corredores de una carrera dada). Esto permite la identificación de la información que afecta a la elección del caballo ganador. CL devuelve la probabilidad de ganar de un caballo, que se calcula en relación con las probabilidades de sus competidores. Esta capacidad explica el porqué de la popularidad del CL en la predicción de las carreras de caballos [31][35][42][55].

El objetivo del modelo de predicción de carreras de caballos CL es predecir un vector con las probabilidades de ganar para la carrera  $j$ , donde la componente  $p_i^j$  representa el modelo de estimación de la probabilidad del caballo  $i$  de ganar la carrera  $j$ , y  $m_j$  denota el número de corredores en la carrera  $j$ . Para lograr esto se define el índice de victorias o *índice de winningness*,  $w_i^j$

$$w_i^j = \beta \cdot x_i^j + \varepsilon_i^j \quad (2.8)$$

Donde:

- $\beta$  es un vector de coeficientes que miden la contribución relativa de las variables de entrada contenidas en el vector.
- $x_i^j$  describe al corredor  $i$  en la carrera  $j$ .
- $\varepsilon_i^j$  es el término de error, representa la información que pasa desapercibida.

Si  $w_i^j$  se define de manera que el caballo con mayor valor de *índice de winningness* gana la carrera  $j$ , entonces se puede demostrar que, si los errores son independientes y distribuidos acordes a la distribución exponencial doble, la probabilidad del caballo  $i$  de ganar la carrera  $j$  es dada por la función CL [43]:

$$p_i^j = \frac{\exp(\beta \cdot x_i^j)}{\sum_{i=1}^{m_j} \exp(\beta \cdot x_i^j)} \quad (2.9)$$

Como dicen Johnson et al. (2006) [35], la elección de este modelo permite que el exponente  $\beta \cdot x_i^j$  se interprete directamente como la habilidad del caballo  $i$ . Los coeficientes del modelo,  $\beta$ , son estimados por medio de un procedimiento de máxima verosimilitud. En particular, dado un conjunto de datos de  $R$  carreras, la probabilidad conjunta  $L = L(\beta)$  es la probabilidad de observación de los resultados, asumiendo que  $p_i^j$  es la de la ecuación (2.9). Por lo tanto

$$\beta \leftarrow \max_{\beta} L(\beta) = \prod_{j=1}^R p_i^j = \prod_{j=1}^R \frac{\exp(\beta \cdot x_i^j)}{\sum_{i=1}^{m_j} \exp(\beta \cdot x_i^j)} \quad (2.10)$$

Donde  $x_i^j$  representa el ganador de la carrera  $j$ .

### 2.1.3.2. Dos etapas previas al modelo propuesto

Estudios previos han demostrado que las *track probabilities* son un buen predictor de los resultados de las carreras [56]. Sin embargo, puede ser perjudicial usarlas junto a variables que describen las habilidades de un caballo en un modelo de predicción [41][46]. En particular, el impacto dominante de las *track probabilities* puede enmascarar a las otras variables e influir de forma negativa en el modelo [48].

Para aliviar este problema y adivinar la verdadera influencia de las variables fundamentales se propone un modelo en dos etapas. En uno de estos modelos, Benter (2003) desarrolló una primera etapa para predecir la posición final de los corredores por medio de una Regresión Lineal Multivariante (MLR) usando solo variables fundamentales. En este paso no se están considerando las *track probabilities*. La estimación de las posiciones finales son interpretadas como una evaluación de las habilidades de los corredores, basadas en actuaciones pasadas, que se pueden observar a través de las variables fundamentales. Posteriormente, estas habilidades se juntan con las *track probabilities* usando CL para estimar las probabilidades de ganar.

$D$  representa una base de datos de  $R$  carreras pasadas con un número de corredores  $M$ .  $D_1$  y  $D_2$  son submuestras disjuntas de  $D$  conteniendo a las carreras  $R_1$  y  $R_2$  con los corredores  $M_1$  y  $M_2$ , respectivamente.  $x_i^j$  es el vector de las variables fundamentales que describen al caballo  $i$  en la carrera  $j$ . Y siendo  $y_i^j$  la posición final. El procedimiento de dos etapas se puede describir:

$$\begin{aligned} f_{MLR}(x) &= \hat{w} \cdot x + \hat{b} \\ \text{Paso 1: } \{\hat{w}, \hat{b}\} &\leftarrow \min_{w, b} \sum_{i=1}^{M_1} (y_i - (w \cdot x_i + b))^2 \end{aligned} \tag{2.11}$$

$$\begin{aligned} \text{Paso 2: } p_i^j &= \frac{\exp(\beta_1 f_{MLR}(x_i^j) + \beta_2 q_i^j)}{\sum_{i=1}^{m_j} \exp(\beta_1 f_{MLR}(x_i^j) + \beta_2 q_i^j)} \\ i &= 1, \dots, M_2; \quad j = 1, \dots, R_2 \end{aligned}$$

Donde:

- $w$  y  $b$  representan la pendiente y la intersección de la función de regresión lineal.
- $\hat{w}$  y  $\hat{b}$  son sus respectivas estimaciones cuadradas, calculadas a través del Paso 1.

Desde que las variables fundamentales son procesadas con el Paso 1 y son resumidas en  $f_{MLR}$ , el segundo paso incorpora el model CL con solo dos entradas de los coeficientes  $\beta_1$  y  $\beta_2$ .

El índice  $j$  se elimina en el primer paso porque la regresión lineal no es capaz de explotar la información relativa al contexto de la carrera. Todos los corredores son considerados independientes y sus posiciones finales son estimadas solamente a través de las respectivas variables fundamentales. Una forma de superar esta restricción y tener la competencia de los corredores en cuenta es reemplazar la regresión lineal del primer paso por una etapa de regresión CL:

$$\begin{aligned} f_{CL}(x_i^j) &= \frac{\exp(\hat{\alpha} \cdot x_i^j)}{\sum_{i=1}^{m_k} \exp(\hat{\alpha} \cdot x_i^j)} \\ \text{Paso 1: } \hat{\alpha} &\leftarrow \max \prod_{i=1}^{M_1} \frac{\exp(\hat{\alpha} \cdot x_i^j)}{\sum_{i=1}^{m_k} \exp(\hat{\alpha} \cdot x_i^j)}, \quad j = 1, \dots, R_1 \end{aligned} \tag{2.12}$$

$$\text{Paso 2: } p_i^j = \frac{\exp(\beta_1 f_{CL}(x_i^j) + \beta_2 q_i^j)}{\sum_{i=1}^{m_j} \exp(\beta_1 f_{CL}(x_i^j) + \beta_2 q_i^j)}, \quad j = 1, \dots, R_2$$

Esta aproximación ha sido aplicada satisfactoriamente en Sung et al. (2005) [42] y en Sung and Johnson (2007) [48] se muestra para superar al modelo de un solo paso.

Edelman (2007) [46] modifica este modelo de dos pasos para superar algunas limitaciones del algoritmo de CL y MLR, respectivamente. En particular, esta técnica deduce un modelo para minimizar el error de los pronósticos en los datos de entrenamiento. Consecuentemente, son propensos a no modelar sólo la estructura si no también el ruido de los datos (sobreajuste de los datos), especialmente si se procesan un gran número de variables fundamentales [49]. Además, son incapaces de explicar las interacciones no lineales entre las variables, a menos que se predefina por el modelador. Edelman ajustó el modelo (2.11) usando SVR [57] en lugar de MLR. Además modificó el procedimiento de SVR original para permitir múltiples términos de intersección. Para

ello, se introduce en (2.11) un término  $b_j$  en lugar de  $b$  para realizar un análisis efectivo de la carrera [46].

Los tres enfoques descritos sólo se diferencian en el primer paso, mientras que la idea general de las variables fundamentales de modelado, y la combinación del primer paso con las *track probabilities* usando el modelo de CL, es idéntico.

MLR/CL se refiere al modelo original de dos etapas (2.11), mientras que CL/CL representa (2.12) y SVR/CL la aproximación de Edelman (2007) [46].

Señalar que el algoritmo SVR sin modificar se usa, en lugar del de Edelman, como punto de referencia en este estudio para obtener una visión más clara en el desempeño competitivo del modelo basado en la clasificación contra el basado en la regresión y, así, tener fiabilidad en la clasificación de los datos.

### **2.1.3.3. Un modelo SVM basado en dos etapas.**

El modelo de predicción de dos etapas, que se desarrolla en este artículo, se basa en el de Edelman (2007) [46]. Se diferencia de este en que en lugar de en la primera fase llevar a cabo una regresión de la posición final de los caballos, el modelo aquí propuesto usa una SVM para obtener un modelo de clasificación para identificar al ganador de una carrera. Esto es motivado por el punto de vista de que, en el contexto de las carreras de caballos, las últimas posiciones no llevan, necesariamente, información importante. Las reglas de las carreras requieren que los jinetes quieran asegurar que el caballo quede en la mejor posición posible, pero hay un incentivo para ellos de conseguirlo cuando no se tienen esperanzas en su victoria. La fiabilidad de las posiciones finales, que constituyen el núcleo del modelo basado en regresión, es cuestionable. Este punto de vista es apoyado por resultados empíricos de Sung and Johnson (2007) [48]. Nuestro modelo se centra en distinguir únicamente ganadores de no ganadores.

El modelo SVM/CL se puede definir de la siguiente forma, donde  $y$  denota un indicador de variable binario de ganar o no ganar, en lugar de una posición final.

$$\begin{aligned}
f_{SVM}(x) &= \left( \sum_{i \in SV} \hat{\alpha}_i y_i \exp(-\gamma \|x_i - x\|^2) \right) + \hat{b} \\
\text{Paso 1: } \{\hat{\alpha}, \hat{b}\} &\leftarrow \max_{\alpha} = \sum_{i=1}^{M_1} \alpha_i \\
&\quad - \frac{1}{2} \sum_{i,k=1}^{M_1} \alpha_i \alpha_k y_i y_k \exp(-\gamma \|x_i - x_k\|^2) \\
&\quad s.t. \sum_{i=1}^{M_1} \alpha_i y_i = 0; \quad 0 \leq \alpha_i \leq C \quad \forall i = 1, \dots, M_1
\end{aligned} \tag{2.13}$$

$$\begin{aligned}
\text{Paso 2: } p_i^j &= \frac{\beta_1 \exp(f_{SVM}(x_i^j))}{\sum_{i=1}^{m_j} \exp(f_{SVM}(x_i^j)) + \beta_2 q_i^j}, \\
i &= 1, \dots, M_2; \quad j = 1, \dots, R_2
\end{aligned}$$

El objetivo del primer paso del modelo es estimar la verosimilitud de un ganador. Por lo tanto, la función signo (2.6) se quita, para obtener una salida continua de SVM. El valor resultante,  $f_{SVM}(x)$ , es proporcional a la distancia de un punto de datos (un caballo) al hiperplano de separación (entre ganadores y perdedores), representando la fiabilidad de un punto (un caballo) perteneciente a una clase particular [49] (si es ganador o perdedor). Si un caballo tiene un alto SVM es más probable que gane.

Como en la mayoría de los modelos anteriores, la información relativa de los caballos que compiten unos contra otros, y el conocimiento previo de cada carrera, se pierde en el paso 1. La SVM es incapaz de explicar la competición dentro de la carrera. Sin embargo, hay que destacar que el objetivo global del modelo de predicción no es maximizar el número de predicciones de ganador correctas, si no maximizar el beneficio. Las apuestas rentables requieren una estimación precisa de la probabilidad de victoria de un caballo. La salida del primer paso sirve solo como un resumen de la habilidad de un caballo (basado en carreras anteriores), mientras que el segundo paso sirve para obtener información sobre la competencia dentro de la carrera.

Lo que se quiere intentar es que la SVM pueda capturar algunos elementos de la competición dentro de la carrera en la primera fase, para mejorar las estimaciones finales. Para ello, se propone una estandarización de la carrera para complementar los

datos. Las variables continuas son estandarizadas comunmente a media cero y desviación típica uno, antes de aplicar el modelo de predicción, para evitar dificultades numéricas con diferentes rangos de valores [58]. Esto se logra restando a la variable su valor medio y dividiendo entre su desviación. Este preprocesado se mejora para tener en cuenta, en cierta medida, la competencia dentro de la carrera. En particular, la estandarización de los datos se realiza de la siguiente manera:

$$\bar{x}_{it}^j = \frac{x_{it}^j - \bar{x}_t^j}{\sigma_t^j} \quad \forall i = 1, \dots, m_j \quad (2.14)$$

Donde:

- $\bar{x}_{it}^j(x_{it}^j)$  denota el nuevo valor del atributo  $t$  del caballo  $i$  en la carrera  $j$ .
- La media  $\bar{x}_t^j$  y la desviación estándar  $\sigma_t^j$  se calculan sobre los corredores de la carrera  $j$ .

Para mostrar la intuición de la normalización de la carrera, consideramos dos carreras de cuatro corredores, con la clase del caballo como entrada única.

	Clase de caballo	Base de datos de normalización	Carrera de normalización
Carrera 1	Caballo 1	20	-0,234
	Caballo 2	40	0,435
	Caballo 3	60	1,104
	Caballo 4	80	1,773
Carrera 2	Caballo 1	1	-0,87
	Caballo 2	3	-0,803
	Caballo 3	5	-0,736
	Caballo 4	7	-0,669

**Tabla 2.1.** Comparación entre estandarización de la carrera y estandarización de la base de datos

La clase de caballo es una medida abstracta de la capacidad del caballo. Los mejores caballos tienen el valor más alto.

## **2.1.4. Evaluación empírica del modelo SVM/CL**

### **2.1.4.1. Datos y variables**

Los datos para los análisis empíricos de este estudio fueron proporcionados por *Raceform Ltd.* y se refieren a las carreras disputadas en el hipódromo de Goodwood en Reino Unido entre mayo de 1995 y agosto de 2000. Este periodo se eligió deliberadamente desde el lugar más grande de intercambio de apuestas en línea, *Betfair*, que se anunció por primera vez en octubre de 2000. La llegada de *Betfair* aumentó la competencia entre los *bookmakers* (los clientes individuales de *Betfair* pueden actuar como corredores de apuestas) y aumentó el número de apuestas profesionales en el mercado. La ventaja de usar datos de antes de septiembre del 2000 es que podemos hacer referencia a resultados empíricos de antes de la llegada de *Betfair* [26]: 200 carreras antes de 1986 [46]: 300 carreras antes de 1995.

Los datos consisten en 556 carreras con 5947 corredores. Las 400 carreras (4296 caballos) realizadas antes de mayo de 1999 son usadas para desarrollar los modelos de predicción, mientras que las 156 restantes se reservan para realizar otras muestras de prueba.

Las variables fundamentales que se usan son las incluidas en Bolton and Chapman (1986) [26] y están proporcionadas en la Tabla 2.2. Estas variables están en dominio público desde nueve años antes de 1995.



Variables independientes	Definición de las variables
Variables relativas al mercado	
$\text{Ln}(q_i^j)$	El logaritmo de las <i>track probabilities</i> normalizadas
Variables fundamentales	
pre_s_ra	La velocidad más alta en la anterior carrera que ha realizado el caballo
avgsr4	La media de la velocidad mayor del caballo en las cuatro carreras anteriores. Con valor 0 si el caballo no tiene carreras pasadas
disavesr	La media de la velocidad más alta en las pasadas carreras, con la misma distancia, hechas por el caballo. Tiene el valor 0 cuando es la primera carrera del caballo
go_avesr	Nivel de velocidad promedio de las carreras pasadas del caballo en ese mismo terreno
draw	Posición en la que sale en la carrera actual
eps	Premios totales ganados (acabando primero, segundo o tercero) hasta la fecha. Y número de carreras en las que pasó
newdis	El valor 1 indica un caballo que ha corrido tres o cuatro de las últimas cuatro carreras con una distancia de menos del 80% de la distancia de la carrera actual. Un 0 para cualquier otra cosa
weight	El peso llevado por el caballo en la carrera actual
win_run	El porcentaje de carreras ganadas por el caballo
jnowin	Número de victorias del jinete hasta la fecha
jwinper	El porcentaje de las victorias del jinete hasta la fecha
jstlmiss	Un 1 indica que se han perdido otras variables relativas al jinete. O en caso contrario

**Tabla 2.2.** Variables usadas para realizar el análisis y su definición

#### 2.1.4.2. Montaje experimental

Las variables continuas del conjunto de datos se normalizan en una base de datos. Los procedimientos de estandarización se han indicado en el punto 3.3.

Una submuestra de 200 carreras (de las 400 de antes de mayo de 1999) son usadas para contruir un clasificador de SVM con *RBF-Kernel* (13). Los parámetros  $C$  y  $\gamma$  son determinados por las medias de cinco veces la validación cruzada [59]. Las 200 carreras se dividen en cinco particiones de igual tamaño y el modelo SVM recursivo se contruye con cuatro particiones, evaluando el resultado sobre la restante. El rendimiento se mide en términos del número de ganadores que se predijo con exactitud. Se

consideran 441 parámetros distintos entre los valores  $\log(C) = \{-3, -2, \dots, 17\}$  y  $\log(\gamma) = \{-20, -19, \dots, 0\}$  [60]. Los valores de los parámetros que conducen a un mayor número de ganadores correctos identificados durante la validación cruzada, se conservan, y se hace un clasificador SVM final con este ajuste basado en 200 carreras.

El modelo de clasificación SVM resultante es usado para marcar las 200 carreras restantes de entrenamiento, proporcionando un índice de habilidad sobre la fuerza relativa de cada caballo y basándose únicamente en las variables fundamentales. Posteriormente, se agrupa el índice de habilidad con las *track probabilities* usando CL en el paso 2 (2.13).

Con el fin de evaluar la rentabilidad del modelo de predicción, se implementa la Estrategia de Apuestas de Kelly [61]. La Estrategia de Kelly identifica cuanto hay que apostar a cada caballo.

Si la metodología propuesta produce un retorno positivo mayor que los otros modelos, se puede decir que el enfoque SVM/CL agrega valor.

#### 2.1.4.3. Evaluación comparativa del modelo de dos etapas.

Las evaluaciones empíricas examinan la eficacia del modelo SVM/CL con la función *Kernel* RBF (que incluyen la base de datos y las variables normalizadas). Por ejemplo, la Estrategia de Kelly, basada en la predicción de la probabilidad de ganar del modelo propuesto, obtiene un retorno del 30,58%.

	Tasa de retorno		$R^2$	Valor t		AUC
	Sin reinversión (%)	Con reinversión (%)		$\beta_1$	$\beta_2$	
SVM/CL	30,58	642,65	0,1323	4,85	10,32	0,762
SVR/CL	17,5	211,55	0,1238	2,86	9,57	0,757
CL/CL	1,74	-16,53	0,1231	2,64	10,53	0,759

**Tabla 2.3.** Comparación empírica de los tres diferentes modelos de dos etapas a través de 156 carreras de exclusión.

Se consideran como puntos de referencia el modelo de dos pasos CL/CL [42] y el procedimiento SVR/CL [46]. Aplicando estas técnicas a las mismas 156 carreras se obtiene una tasa de rendimiento de 1,74% y 17,50% respectivamente (ver tabla 2.3). Comparaciones parecidas se pueden observar cuando se permite la reinversión de las ganancias. Los dos métodos de apoyo basados en los vectores, SVM/CL y SVR/CL, producen un significativo aumento en la riqueza (642,65% y 211,55%, respectivamente), mientras que la riqueza descende en un 16,53% cuando usamos el modelo CL/CL.

El método propuesto SVM/CL supera a sus dos competidores, proporcionando significativamente beneficios más altos usando el Criterio de Kelly con y sin reinversión. Los resultados de usar el metodo Kelly sin reinversión son un indicador más fiable del éxito del modelo relativo, en cuanto a que los beneficios recogidos con reinversión pueden surgir de la suerte del orden en el que se producen los ganadores y perdedores. Como consecuencia, en los análisis posteriores nos centramos en los resultados sin reinversión.

Tambien podría argumentarse que confiar solamente en la rentabilidad de un modelo particular pasa por alto otros indicadores clave de rendimiento de un modelo. Por lo tanto, la Tabla 2.3 resume los indicadores de rendimiento adicionales para facilitar la comparación posterior de los tres métodos.

El factor  $R^2$  representa el tanto por ciento de la variación de la rentabilidad de un modelo. De esta forma el del modelo SVM/CL (0,132) excede del de sus dos competidores, indicando que las probabilidades generadas contienen información más útil en las variables fundamentales. Esto se confirma cuando examinamos los valores de la t-estadística de  $B_1$ , el coeficiente CL asociado con la salida de la etapa 1 (ecuaciones (2.11)-(2.13)). El índice de la capacidad obtenida mediante el procesamiento de las variables fundamentales en la primera etapa con una SVM tiene el valor  $t$  más alto y puede, por consiguiente, ser considerado como el más informativo.

	Tasa de retorno		R <sup>2</sup>	Valor t		AUC
	Sin reinversión (%)	Con reinversión (%)		$\beta_1$	$\beta_2$	
Modelo de referencia propuesto SVM/CL	30,58	642,65	0,1323	4,85	10,32	0,762
Modelo de una etapa CL	-0,46	-48,37	-	-	-	0,737
SVM	-22,57	-100,00	-	-	-	0,761
Modelo de dos etapas empleando distintas variables de entrada						
Solo Base de datos normalizada	5,86	-116,33	0,12	1,14	11,87	0,756
Solo Normalización por carreras	3,46	-0,04	0,123	3,26	10,72	0,756
Modelo de dos etapas SVM con <i>Kernel</i> lineal						
SVM <sub>lineal</sub> /CL	7,35	120,73	0,1193	0,188	13,2	0,756

**Tabla 2.4.** Comparación empírica de otros modelos con las 156 carreras de exclusión.

El poder discriminativo de un modelo en términos de su área bajo una curva de características de recepción de funcionamiento (AUC) se considera también un indicador de rendimiento (actuación, desempeño) [62]. El AUC es una medida popular para la evaluación de los clasificadores. Para esta aplicación, representa la probabilidad de que un modelo asigne una mayor posibilidad de ganar a un caballo ganador que a uno perdedor. Los valores prácticos de AUC van de 0,5 a 1 donde los valores más altos representan un poder discriminativo más alto. Las diferencias de rendimiento en términos de AUC entre los tres modelos son menores (ver tabla 2.4). Estos resultados sugieren que aunque SVM/CL es solo ligeramente mejor en términos de identificar ganadores, sobre todo destaca en producir las probabilidades de ganadores más exactas y de esta manera obtener más beneficios.

La mayor diferencia entre la aproximación propuesta aquí y el trabajo de Edelman (2007) [46] es el uso del rango de datos ordenados en la construcción del modelo. Los resultados sugieren que la clasificación puede ser más fiable que la regresión para los datos de carreras de caballos. Este punto de vista se apoya en la realización de una prueba formal para examinar si las carreras que excluyen al ganador final sigue la

misma distribución, que las carreras en las que se tiene en cuenta al ganador [63][64]. La prueba estadística correspondiente ( $\chi^2_{13} = 21,20$ ) indica que esta hipótesis debería ser rechazada en un 7%, proporcionando una prueba más de la falta de fiabilidad del orden de las posiciones finales.

#### **2.1.4.4. Examinando el origen del beneficio**

A pesar del atractivo rendimiento del modelo SVM/CL, los modelos SVM son métodos que no nos proporcionan explicación de las relaciones entre los datos. De esta forma, se requieren simulaciones adicionales para arrojar luz sobre el origen del beneficio con el método de apuestas basadas en el Modelo de Kelly. Además, el hecho de que usemos la clasificación SVM en la primera etapa, vamos a tener tres factores que afectan el rendimiento del modelo; concretamente, ejecutar un modelado en dos etapas que no emplea el uso de las *track probabilities*, emplear un enfoque de pre-procesamiento de un conjunto de datos especiales para capturar alguna información sobre la competencia dentro de la carrera, y usar un modelo no lineal (ejemplo función *Kernel* RBF), para distinguir entre los ganadores y los perdedores. Los resultados de estos experimentos se resumen en la Tabla 2.4.

##### *2.1.4.4.1. Comparación entre modelos de una y de dos etapas.*

Respecto al enfoque del modelo de dos etapas, Sung and Johnson (2007) [48] muestran que es superior a modelos de una etapa cuando usamos CL. Sus resultados se confirman con los datos empleados en este estudio: un modelo de una etapa CL produce una pérdida de 0,46% durante las carreras cuando se aplica la Estrategia Kelly sin reinversión. Sin embargo, el modelo de una etapa SVM funciona mucho peor, teniendo una pérdida de 22,57% durante las mismas carreras (Tabla 2.4). Las probabilidades ganadoras producidas por los modelos de una sola etapa CL o SVM son significativamente inferiores a los modelos respectivos de dos etapas.

##### *2.1.4.4.2. Comparación de modelos lineales y no lineales.*

El modelo propuesto SVM/CL incorpora una función de *Kernel* RBF para tener en cuenta las relaciones no lineales entre las variables fundamentales. El rendimiento superior del SVM/CL sobre el modelo CL/CL, que acoge solo las relaciones lineales,

indica que la relación entre variables independientes y los resultados de la carrera no es lineal. Sin embargo, para obtener una visión más clara de este tema vamos a ver el rendimiento de un modelo SVM/CL con un *Kernel* lineal (ultima fila de la tabla 4). Una simulación de apuestas de este modelo sobre las 156 revela que el modelo lineal SVM/CL produce un rendimiento, retorno inferior (7,35% sin reinversión) al producido por el modelo no lineal SVM/CL (30,58%). Similarmente, los otros indicadores de rendimiento demuestran la superioridad del modelo no lineal. Dado que la función de *Kernel* es la única diferencia entre estos dos modelos, se puede concluir que las relaciones no lineales que existen entre variables independientes deberían ser tomadas en cuenta cuando modelamos el resultado de la carrera.

#### 2.1.4.5. Debate

Es importante recordar que las únicas variables usadas en este estudio son aquellas incluidas en el artículo de Bolton y Chapman (1986) [26]. Se esperaba que como habían sido de dominio público durante muchos años, el publico tuviera la información completa de estas variables. A pesar de esto, el modelo CL/CL se las arregla para generar un pequeño beneficio sobre la muestra de las 156 carreras (1,74%) si las ganancias no se reinvierten.

Ambos métodos logran una significativa mejoría en términos de beneficio (SVM/CL: 30,58% y SVR/CL: 17,50%). Este resultado confirma los resultados anteriores de Edelman (2007) [46] que incorporan un modelo de dos pasos que permite incluir la información de las variables fundamentales, el cual no ha sido todavía tomado en cuenta por los apostantes. Además, el enfoque propuesto basado en la clasificación, SVM/CL tiene importantes mejoras adicionales sobre un enfoque SVR/CL. De esta forma se confirma la sospecha de que no podemos realizar de forma fiable un pronóstico basado en la regresión de las posiciones finales.

Sin embargo, usar clasificación tiene algunos inconvenientes teóricos. En particular, la variable binaria de objetivo ganar/perder no es independiente en cada caballo. Una formulación multinomial SVM [65] podría ser considerada como una alternativa. Esto podría implicar definir la posición final de un corredor como una variable de objetivo discreto y construir un modelo SVM que distingue los caballos que

terminan primero, segundo etc. Sin embargo, tal enfoque tiene grandes inconvenientes. En particular, las carreras pueden incluir un gran número de corredores y se necesita una clase para cada posible posición final.

Un tema clave en predecir el resultado de los eventos de carreras está dentro de la competición de cada carrera. Desde un punto de vista metodológico, el único procedimiento actualmente disponible capaz de complacer las relaciones entre los corredores es el CL. Por otra parte, los resultados de Benter (1994) [41], Edelman (2007) [46] y aquellos presentados aquí indican que las técnicas ordinarias de predicción que consideran cada ejemplo como independiente se adaptan bien, si se combinan con CL dentro de una segunda etapa. De hecho, la combinación con CL es esencial como se demuestra en la sección 4.4.1. Consecuentemente, se pueden esperar mejoras adicionales cuando se hace posible también modelar relaciones entre los caballos en una carrera en la etapa uno.

Otros modelos, como los ordinales SVM [66] o el de regresión logística de *Kernel* [67], podrían ser considerados, ya que siguen las mismas ideas que el modelo SVM, pero encarnan diferentes tipos de funciones de pérdida.

Sin embargo, todos estos procedimientos miden la pérdida sobre ejemplos individuales y agregan estos valores para formar una medida general del error empírico. Este es el paso donde las dependencias entre ejemplos se pierde. Los recientes avances en el campo de los SVM estructurales podrían ofrecer una alternativa al permitir funciones de pérdida más complejas que no restringen los ejemplos individuales. Por ejemplo, Joachims (2005) [68] desarrolla una SVM que optimiza la AUC directamente. Esta técnica parece ser una buena candidata para la primera etapa del modelado de carreras de caballos en futuros trabajos.

### **2.1.5. Conclusión**

Se propone una metodología de dos etapas para predecir resultados de eventos competitivos. Este modelo difiere de los otros de dos etapas en que considera la clasificación en lugar de la regresión en la primera etapa para evitar problemas con el rango poco fiable en el orden de los caballos. En cambio, lleva a cabo un procedimiento de estandarización de los datos dentro de la carrera para dar al modelo basado en SVM

información dentro de la competición de las carreras. Los resultados empíricos han demostrado la efectividad del modelo en ofrecer mejoras considerables de precisión sobre alternativas de la competencia.

Los resultados indican que, aunque los modelos de apuestas de carreras de caballos usan variables fundamentales similares que han sido de dominio público durante muchos años [26][41][44], el público apostante todavía no descifra completamente la información contenida en estas variables. Esto refleja la complejidad de la relación entre las variables fundamentales y el resultado de la carrera, que en vista de los resultados observados, es proclive a incluir iteraciones no lineales y podría mantenerse oculto a aquellos que pueblan los mercados. En el trabajo futuro, las técnicas de extraer reglas de clasificadores SVM, de entrenamiento [69] podría ser aplicado para explorar la naturaleza de las relaciones entre las variables y mejorar la comprensión de la información que los individuos de estos mercados no consiguen descifrar. Además, sería interesante llevar a cabo otros experimentos adicionales usando datos después de la llegada de *Betfair*. Comparando tales resultados a los presentados aquí podría ayudar a cuantificar el grado en que Internet ha cambiado el enfoque de los mercados de apuestas de carreras de caballos.

## 2.2. Modelo aplicado

En este apartado aplicaremos el modelo descrito antes a nuestro caso particular, de forma que, primero explicaremos como hemos hallado la base de datos de resultados de carreras de caballos, y posteriormente calcularemos las posibilidades que tiene cada caballo de resultar ganador en una determinada carrera, así como el dinero que apostaremos por cada caballo, y la ganancia total que obtendremos con esas apuestas.

### 2.2.1. Página web Formstar

Ya hemos hablado de algunas de las casas de apuestas que existen para carreras de caballos.

En nuestro caso nos hemos basado en datos obtenidos en la página web de Formstar [1]. Es una página de Sudáfrica que recoge carreras de varios tipos corridas en












distintos hipódromos. La elección se realizó tras el estudio de varias páginas web que recogían información acerca de distintos caballos. En esta página encontramos variables parecidas a las que se nos presentaban en el artículo anterior, y además contenía numerosa información acerca de las características de los caballos, de los jinetes o los entrenadores.

Vamos a describir la estructura de la página en cuestión, ya que a simple vista puede parecer un poco complicado de entender.

1. Página principal, **Home**, donde nos dan algunos datos relativos a Formstar, noticias, etc.
2. Pestaña **Meetings** donde se encuentra la información que más nos va a interesar. Nada más introducimos en ella nos salen todos los días de carreras que se han corrido y algunos que se correrán en los días próximos.

En todas las ventanas de esta pestaña nos aparecerán algunas (o todas) de las siguientes opciones:

-  → Nos sirve para ir a la ventana de Races, estemos donde estemos.
-  → Sirve para ir viendo todas las carreras corridas anteriormente.
-  → Se utiliza para ir viendo todas las carreras corridas con posterioridad.
-  → Sirve para buscar la carrera que quieras sin necesidad de ir bajando manualmente.
-  → Nos salen las estadísticas de apuestas que han sido las mejores a lo largo de los años (desde 2003). Los estudios son del primero, segundo, tercero y cuarto.

-  → Es una de las opciones más importantes, ya que te explica uno a uno todas las columnas que contienen cada página, los significados de las palabras que se emplean, los números, las iniciales, etc.
-  → Sirve para ir a la ventana de *field*. Esta opción se encuentra en Forecast y en Result.
-  → Para pasar a la ventana Result.
-  → Ir a la ventana Forecast.

Nos dan información como:

- Fecha de la carrera.
- Donde se realiza la carrera.
- Tipo de terreno.
- Si ya se ha corrido (Result) o si aún está pendiente de correrse (Forecast).
- El número de carrera.
- El número que lleva el caballo en el dorsal.
- Nombre del caballo por el que se ha hecho la mayor apuesta.
- Las *odds* de ese caballo.
- Si ya se ha corrido la carrera, la posición en la que ha quedado finalmente el caballo que se pensaba que iba a ganar.
- Y por último el precio de salida, es decir la cuota a la que se pagará por ese caballo cuando empiece la carrera.

A continuación podemos meternos en un día de carrera y nos encontramos con todas las categorías que se han corrido. Es decir, en un día de carrera se han podido realizar varias carreras de distintas o de la misma categoría (a distintas horas y con distintos caballos). Estamos en la ventana **Races**.

En este caso las distintas columnas nos dan información de:

- Número de carrera.
- Tipo de carrera.
- Longitud del hipódromo.
- Apuestas realizadas sobre esa carrera.
- Fuerza que tiene el caballo (es un factor calculado por la propia página).
- Número de caballos que participan.
- Nombre del corredor ganador (en caso de que ya se haya corrido).
- Tiempo que ha tardado en realizar toda la carrera el ganador.
- Número de dorsal del caballo que según la propia página tiene más probabilidad de ganar.
- Nombre del caballo anterior.
- Probabilidades calculadas por FormStar.
- Posición final de ese caballo.
- El precio de inicio actual.

La siguiente ventana que localizamos (al pinchar en el tipo de carrera que quieras) es la de **Forecast**. En esta página vamos a tener más información acerca de los caballos que han realizado la carrera

- Posición predicha de cada caballo.

- Posición en la que realmente acaba (si se conoce).
- Número de dorsal del corredor.
- Últimos cinco resultados del corredor. Siendo 1 = primero, 2 = segundo,..., 0 = décimo o peor.
- Nombre del caballo. Si el caballo no ha sido criado localmente, el origen está entre paréntesis.
- Si lleva *blinkers*.
- Posición de salida.
- *Ranking* de desventaja de cada caballo. Cuanto mayor es mejor.
- Nombre del jinete.
- Calificación del jinete.
- Masa total que lleva el caballo.
- Rango realizado por el oficial de hándicap según el mérito del caballo.
- Entrenador del caballo.
- Calificación del entrenador.
- Calificación de la combinación del entrenador y el jinete.
- Clasificación de FormStar.
- A cuantos metros está pronosticado que quede cada caballo del ganador.
- Las probabilidades calculadas por FormStar de ganar el corredor en esa carrera.
- Precio de salida del caballo. F denota al favorito (solo si se conoce el resultado) .

- Carreras corridas por el caballo.
- Carreras ganadas por el caballo.
- Carreras en las que el caballo terminó colocado.
- Apuestas ganadas (aproximadamente) por el caballo.
- Para miembros *Premium* está la opción de añadir un comentario sobre la carrera.

También se nos da la opción de ver más datos de los caballos pinchando sobre el que queramos. Los datos que nos ofrecen en la pestaña **Form** son:

- Fecha.
- Hipódromo. Nos vamos a encontrar con unas siglas cuyo significado es:
  - Arln: *Arlington*.
  - Bftn: *Bloemfontein*.
  - Clwd: *Clairwood Park*.
  - Dbvl: *Durbanville*.
  - Fair: *Fairview*.
  - Flmg: *Flamingo Park*.
  - Gosf: *Gosforth Park*.
  - Grey: *Greyville*.
  - Kenw: *Kenilworth*.
  - Kimb: *Kimberley*.
  - Maur: *Mauritius*.

- Nmkt: *Newmarket*.
- Rand: *Randjiesfontein*.
- Scot: *Scottsville*.
- TrfN: *Turffontein (New course - inside track)*.
- Turf: *Turffontein*.
- VSnd: *The Vaal (Sand)*.
- Vaal: *The Vaal*.
- Zimb: *Zimbabwe (Borrowdale)*.
- Condiciones del terreno, cuyo significado es:
  - F: firme.
  - G: bueno.
  - H: denso
  - R: firme.
  - S: flexible/suave.
  - Y: blando.
- Tipo de carrera:
  - AdvP: *Advanced Plate*.
  - AlwP: *Allowance Plate*.
  - Cndp: *Conditions Plate*.
  - Fhcp: *Durbanville*.
  - FSt2: *2yo Feature Stakes*.

- FSt3: *3yo Feature Stakes*.
- FStk: *Feature Stakes*.
- GrdP: *Graduation Plate*.
- HcpA: *A Division Handicap (merit rated 99 and over)*.
- HcpB: *B Division Handicap (merit rated 91 to 98)*.
- HcpC: *C Division Handicap (merit rated 83 to 90)*.
- HcpD: *D Division Handicap (merit rated 75 to 82)*.
- HcpE: *E Division Handicap (merit rated up to 74)*.
- JuvP: *Juvenile Plate*.
- MJvP: *Maiden Juvenile Plate*.
- MdnP: *Maiden Plate*.
- NovP: *Novice Plate*.
- PrgP: *Progress Plate*.
- SupP: *Superior Plate*.
- Potencia de cada corredor en la carrera.
- Distancia.
- Si el caballo lleva *blinkers*.
- Posición de salida.
- Nombre del jinete.
- Peso llevado por el caballo en kg.
- Rango realizado por el oficial de handicap según el mérito del caballo.

- Entrenador.
- Calificación de los corredores que tienen menos ventaja según Formstar (antes de la carrera).
- Número de corredores.
- Posición final
- A cuanto quedan los corredores del primero, si es el primero, a cuanto queda del segundo.
- Corredor con desventaja en la carrera segun FormStar.
- Tiempo en completar la carrera (en segundos).
- Clasificación de velocidad de FormStar para los corredores.
- Precio de salida del corredor. F es el favorito.
- El número de carreras finalizadas por el caballo.
- Número de carreras ganadas.
- Número de veces que ha quedado colocado.
- Apuestas ganadas (aproximadamente) por el caballo.

A partir de aquí vamos a pasar a describir muy por encima el resto de pestañas que nos encontramos en la página web de Formstar, pero que en este Proyecto no las hemos tenido en cuenta para crear la base de datos.

3. La siguiente pestaña que tenemos es la de **Horses** (caballos). Lo que nos encontramos en ella es un cuadro resumen con una clasificación de los veinte mejores caballos. Se nos da información como la situación dentro del *ranking*, posición que ha quedado en las últimas cinco carreras, nombre del caballo, descripción del caballo, carreras corridas, carreras ganadas, porcentaje de carreras ganadas, número de veces que ha terminado colocado, porcentaje de las veces que



ha quedado colocado, la calificación de *Formstar*, dinero ganado, si el caballo sigue activo o no y la última carrera realizada por el mismo.

4. A continuación tenemos la pestaña ***Connections*** (relaciones) con información de ***Trainers*** (entrenadores) y ***Jockeys*** (jinetes).

Los datos que nos aportan de los jinetes y de los entrenadores son los mismos. La posición que ocupa dentro de la clasificación, el nombre del mismo, carreras corridas, carreras que ha terminado primero, segundo, tercero y cuarto (cada una en una columna distinta), porcentaje de carreras ganadas, porcentaje de carreras en las que ha terminado colocado, apuestas ganadas y clasificación de *Formstar* que se calcula en el caso de los jinetes haciendo un promedio ponderado exponencial del número de puntos de los caballos montados por el jinete y en el caso de los entrenadores de los caballos que han sido entrenados por él.

5. Las siguientes pestañas: ***Subscriptions***, ***Support***, ***Info*** y ***Login*** no aportan ningún dato relativo que pueda ser tenido en cuenta para la realización de la base de datos.

### 2.2.2. Base de datos

Uno de los principales objetivos de este Proyecto Fin de Carrera era crear una base de datos con los resultados de carreras pasadas. Esto es así porque, al contrario que en otros deportes, en el que nos atañe, no tenemos los resultados de forma fácilmente accesible. Por lo tanto teníamos que obtenerlos a partir de páginas web que fueran poniendo a disposición del usuario la información según se fueran realizando las carreras, y, a poder ser, que también contara con un histórico de carreras anteriores para abarcar una mayor cantidad de datos.

Para crear la base de datos ha sido necesario realizar un programa que leyera de la página web escogida (explicada en la sección anterior).

La página web se trata como un fichero, por lo que hay que realizar la apertura del mismo para empezar a ser leído, además de cerrarlo cuando terminemos.

Para que fuera más sencilla su lectura hemos quitado las negritas (había palabras que estaban en negrita y otras no, sin seguir ninguna lógica), y además hemos realizado

una justificación a izquierdas de todas las columnas. Por lo tanto, todos los datos que nos interesan van a estar entre el código html: `<td class==left>` y `</td>`. De esta forma iremos buscando ese código, y una vez que estemos colocados en el sitio exacto, cogeremos los datos que nos van a ir interesando.

No hemos encontrado ninguna página que tuviera exactamente todas las variables que nos han presentado en el artículo, sin embargo, podemos utilizar otras con los mismos resultados. La pestaña, de las descritas antes, que vamos a emplear va a ser *Forecast*.

Para ir guardando las variables, o referencias a las mismas, que van a formar parte de nuestra base de datos, vamos a realizar una matriz numérica.

- La primera variable del archivo que nos va a interesar va a ser la que está en la segunda columna, que es la posición final en la que ha terminado cada caballo. Como estamos usando carreras pasadas para realizar la base de datos, esta posición ya va a ser conocida.

Aquí lo único que vamos a diferenciar es si un caballo ha quedado en primer lugar o no, por lo tanto vamos a crear una columna con un 1 en la posición del ganador y un 0 en el resto. Para ello lo que hacemos es leer el valor, si es igual a 1 asignamos a la columna de una matriz numérica, a la que llamaremos resultado, un 1, si es distinto de 1 le ponemos un 0. De la siguiente forma:

	Ganador
Caballo 1	0
Caballo 2	0
Caballo 3	0
Caballo 4	1
Caballo 5	0
Caballo 6	0
Caballo 7	0
Caballo 8	0
Caballo 9	0

**Tabla 2.5.** Ejemplo matriz del caballo ganador

- Otro dato que vamos a tener en cuenta van a ser las últimas cinco posiciones en las que han quedado los caballos que realizan la carrera. Esta información nos la proporciona la cuarta columna del archivo (*Form*). Cada posición la vamos a ir guardando en una columna distinta de la matriz numérica. Es decir, la primera columna contendrá información sobre la posición en la que quedó un caballo determinado hace cinco carreras, la segunda hace cuatro, y así sucesivamente. Cuando un caballo no ha realizado las cinco carreras vamos a poner en esa columna el valor 11. Además, cuando ha quedado en décima posición o peor, la rellenaremos con valor 10.

Para asignarlo, lo que tenemos que hacer es comprobar la longitud de lo que estamos leyendo, con esto sabremos cuantas carreras anteriores ha realizado, y así sabremos si no ha realizado 5 carreras anteriormente, de forma que esas posiciones las rellenaremos con valor 11. También tendremos que ir asegurándonos de si lo que estamos analizando es un 0 u otro valor, de forma que si es un 0 le asignaremos un 10 a esa posición de la matriz numérica y si es otro valor lo rellenaremos con él mismo.

Ponemos un ejemplo: Suponemos que lo que estamos leyendo es 503. Como vemos, la longitud es 3. De forma que las dos primeras columnas de la matriz numérica las rellenaremos con un 11. El siguiente es un 5, no es ni 0 ni vacío, por lo tanto lo asignamos directamente. Como el posterior es un 0, en ese caso pondremos un 10 en esa posición. El 3 lo podremos poner directamente también. Así quedaría:

Columna 2	Columna 3	Columna 4	Columna 5	Columna 6
11	11	5	10	3

**Tabla 2.6.** Ejemplo de relleno de columnas con variable *Form*

- Las siguientes seis columnas de la matriz numérica las vamos a ir rellenando con el *ranking* de desventaja de cada caballo, la calificación del jinete, el peso total que lleva el caballo en kilogramos, la clasificación del mérito del caballos según *Formstar*, la calificación del entrenador y la combinación de las calificaciones del jinete y del entrenador. Estos datos los asignamos directamente a nuestra matriz numérica, ya que en todos los casos los datos se nos dan en forma de números.

- El siguiente dato que nos interesa es la columna dieciséis del fichero, esto es la clasificación del caballo según *Formstar*. Como es un número lo puedo asignar directamente a la matriz numérica. Si en algún caso nos encontramos que en una posición está la variable *poor*, en la posición correspondiente de la matriz numérica pondremos el valor -2.
- En la columna diecisiete está el pronóstico de a cuantos metros ha quedado el caballo del primero, que también se trata de un número por lo que no tenemos que hacer nada para asignarlo.
- Luego leemos del fichero la columna dieciocho y la diecinueve, aquí encontramos las *odds* del caballo y su precio de salida, respectivamente. Los dos datos nos los dan en formato “a-b”. Lo que queremos es la relación entre ellas, por lo que vamos a leer desde el principio hasta el guión y lo guardamos en una variable auxiliar, y posteriormente desde el guión hasta el final, guardándolo en otra variable. En algunos casos en lugar de “-” nos encontramos con un “/”, es decir, “a/b”. En este caso buscamos la barra y la reemplazamos por el guión, así procedemos de la misma manera que si desde el principio el formato fuera “a-b”. Una vez obtenidos los dos valores dividimos el primero entre el segundo, es decir, en el ejemplo a dividido entre b.

Por ejemplo:

Odds	aux1	aux2	Columna 15
7-2	7	2	$7/2 = 3.5$

**Tabla 2.7.** Ejemplo de relleno de columnas con las cuotas

Otra cosa a tener en cuenta, es que hay caballos en los que junto a las *odds* y a los SP tenemos una “F” denotando al favorito. Para ella vamos a hacer, a continuación, una columna cuyos valores serán 0 (si no hay F) y 1 (si hay F).

Hay ocasiones en que nos encontramos, que por algún error, estas dos columnas no tienen ningún valor, es decir, esas celdas están vacías. En ese caso lo que hemos hecho es asignarle la palabra “NADA” a la posición correspondiente de la página html en cuestión, es decir:

<td class==left></td> → <td class==left>NADA</td>

Por lo tanto, también tenemos que hacer la comprobación de si lo que estamos leyendo es la palabra “NADA” y si es así, ponemos un 0 en las cuatro posiciones de nuestra matriz (tanto la del valor de *odds* como la de SP, así como las F correspondientes)

- A continuación, volvemos a tener tres valores numéricos, que son R (carreras corridas por el caballo), W (carreras ganadas) y P (número de veces que ha quedado colocado).
- Y, por último, en la columna veintitrés del fichero están las apuestas ganadas por el caballo. En este caso los datos se nos dan en formato decimal con coma y nosotros para poderlo asignar a la matriz numérica lo necesitamos en formato decimal con punto. Para ello reemplazamos la una por la otra y lo asignamos con formato *double*, para los números elevados.

Ahora vamos a proceder a leer otros datos que nos interesan del fichero, para tener información importante de las características de los caballos que realizan la carrera, pero que se encuentran en el título de la página. Estos son el terreno, la distancia y el factor R.

Para ello primero tenemos que colocarnos en la posición adecuada del fichero. Esto lo hacemos buscando la primera palabra de la frase y colocándonos justo donde acaba. También aquí todo lo que leamos va a estar entre los códigos html <td class==left> y </td>.

- Lo primero que vamos a leer va a ser el tipo de terreno. Para ello vamos a buscar los guiones de la frase en la que está inmerso, que es en la cabecera de la página web, detrás de *Meeting*.

Por ejemplo: Meeting: Sat 12-Feb-2011 - Kenilworth - Going: Good

Como lo que nos interesa está a partir del cuarto guión, nos situamos ahí. Desde el cuarto guión hasta el final siempre estará la palabra *Going*: seguida del terreno, por lo que no vamos a leer desde justo el guión, si no unas cuantas posiciones más

tardías (para ser concretos siete posiciones) hasta el final. Para asignarlo a la matriz numérica hemos hecho una correspondencia del siguiente modo:

Tipo de terreno	Correspondencia numérica
Good	1
Firm	2
Heavy	3
Hard	4
Soft	5
Yield	6

**Tabla 2.8.** Correspondencia numérica con el terreno

Los dos siguientes datos se encuentran en la misma frase, detrás de la palabra *Race*, por lo que tenemos que realizar la búsqueda de esa palabra y de los guiones sólo una vez.

En estos dos últimos casos vamos a tener que distinguir entre dos situaciones:

- Que la carrera sea de las *Division*. Ejemplo: Race: 6 - D Div Handicap (MR75-82) - 1000m - R 57,040
- El resto de carreras. Ejemplo: Race: 4 - Maiden Plate - 2500m - R 53,360

Esto es así, por que nuestra forma de buscar los datos que nos interesan, va a ser buscando el guión a partir del cual se encuentra el mismo, y en estos dos casos el número de guiones es distinto. Teniendo cuatro guiones en el caso de las carreras *Division*, ya que hay un guión inmerso en el tipo de carrera, y tres guiones en el resto.

- La distancia de la carrera es la que se encuentra en primer lugar. Por lo tanto lo que buscamos estará entre el guión tercero y cuarto en las carreras *Division*, y en el resto la información estará entre el segundo y el tercer guión.
- Para el factor R, tendremos que distinguirlo también. En el primer caso estará entre el cuarto guión y el final. Y en el segundo estará entre el tercero y el final.

Modo resumen:

- Carrera de las *Division*:
  - Distancia: entre guiones tercero y cuarto.
  - Factor R: entre cuarto guión y final.
- Resto de carreras:
  - Distancia: entre guiones segundo y tercero.
  - Factor R: entre tercer guión y final.

Adjuntamos una tabla como resumen de lo anterior.

Número columna	Datos
1	Posición caballo hace 5 carreras 10: Posición $\geq 10$ 11: Carrera no corrida
2	Posición caballo hace 4 carreras 10: Posición $\geq 10$ 11: Carrera no corrida
3	Posición caballo hace 3 carreras 10: Posición $\geq 10$ 11: Carrera no corrida
4	Posición caballo hace 2 carreras 10: Posición $\geq 10$ 11: Carrera no corrida
5	Posición caballo carrera anterior 10: Posición $\geq 10$ 11: Carrera no corrida
6	Drtg
7	Calificación jinete

8	Peso (Kg)
9	MR
10	Calificación entrenador
11	Combinación jinete-entrenador
12	Fcast -2: Poor
13	Metros del primero
14	<i>Odds</i>
15	Caballo Favorito 1: Favorito 0: No Favorito
16	SP
17	Caballo Favorito 1: Favorito 0: No Favorito
18	Carreras corridas
19	Carreras ganadas
20	Número de veces colocado
21	Apuestas ganadas
22	Terreno 1: <i>Good</i> 2: <i>Firm</i> 3: <i>Heavy</i> 4: <i>Hard</i> 5: <i>Soft</i> 6: <i>Yield</i>
23	Distancia
24	Factor R

Tabla 2.9. Resumen de los datos de la matriz numérica



A continuación vamos a poner un ejemplo de cómo quedaría la matriz numérica en una carrera en particular.

	Columna 1	Columna 2	Columna 3	Columna 4	Columna 5	Columna 6	Columna 7	Columna 8	Columna 9	Columna 10	Columna 11	Columna 12	Columna 13
Caballo 1	2	11	11	11	11	25	-27	60	0	25	-30	1213	0
Caballo 2	7	6	9	2	11	-3	25	60	68	39	4	1049	1,2
Caballo 3	10	3	11	11	11	-165	-27	60	0	62	-26	1032	1,33
Caballo 4	10	8	3	4	11	-81	-20	60	70	-62	-25	998	1,58
Caballo 5	4	6	4	11	11	-78	19	60	75	-62	-14	933	2,06
Caballo 6	10	10	7	9	10	-52	64	60	55	-18	-108	700	3,77
Caballo 7	10	7	9	8	10	-112	18	60	66	-79	-37	628	4,3
Caballo 8	2	10	4	10	10	-102	-77	60	61	-79	-168	567	4,75
Caballo 9	10	11	11	11	11	-55	38	60	0	-192	-18	513	5,14

Tabla 2.10. Ejemplo de matriz numérica con una carrera particular (columnas 1 a 13)

	Columna 14	Columna 15	Columna 16	Columna 17	Columna 18	Columna 19	Columna 20	Columna 21	Columna 22	Columna 23	Columna 24
Caballo 1	2,5	0	1,4	1	1	0	1	10,672	1	1200	53,36
Caballo 2	4,5	0	3,5	0	4	0	1	10,672	1	1200	53,36
Caballo 3	5	0	6,5	0	2	0	1	5,336	1	1200	53,36
Caballo 4	5,5	0	6,5	0	4	0	2	8,004	1	1200	53,36
Caballo 5	7	0	5,5	0	3	0	2	5,918	1	1200	53,36
Caballo 6	20	0	22	0	15	0	5	43,5	1	1200	53,36
Caballo 7	25	0	25	0	9	0	1	11,6	1	1200	53,36
Caballo 8	40	0	50	0	29	0	16	108,518	1	1200	53,36
Caballo 9	66	0	14	0	1	0	0	0	1	1200	53,36

Tabla 2.10. Ejemplo de matriz numérica con una carrera particular (columnas 14 a 24)

### 2.2.3. Inferencia

El principal objetivo de este apartado es conseguir, a partir de nuestra base de datos calculada en el punto anterior, las probabilidades que tiene cada caballo de ganar en cada carrera.

Para realizar una prueba del modelo lo que hemos hecho ha sido contar con todas las carreras de categoría E Division cuyo hipódromo es de 1400 metros de longitud. De esta forma tenemos 40 carreras, cada una con un número determinado de caballos.

Lo primero que tenemos que hacer es dividir el conjunto de datos de las 40 carreras en los siguientes subconjuntos:

- Datos de entrenamiento. Son aquellos que se usan para realizar un entrenamiento de la máquina con el fin de que se ajuste mejor a nuestras variables.
- Datos de test o prueba. Estos datos serán los que nos permitirán validar como de bueno es el análisis en otro subconjunto no incluido en el anterior.

Para realizar la división del conjunto de datos hemos empleado dos técnicas que se basan en la validación cruzada, también conocida como *cross-validation*. Es una técnica que se utiliza para evaluar los resultados de un análisis estadístico, y garantizar que son independientes de la partición entre datos de entrenamiento y de test. Se emplea para estimar cómo de preciso es un modelo.

La validación cruzada [72] es una forma de evaluar un modelo con un hipotético conjunto de datos de prueba, cuando no disponemos del conjunto en sí.

Existen muchos tipos de validación cruzada, sin embargo los que vamos a usar son:

1. **Validación cruzada de  $K$  iteraciones (CV).** En este tipo de validación los datos de muestra se dividen en  $K$  subconjuntos, uno de los cuales se usa como datos de test, mientras que el resto ( $K-1$ ) se usan como datos de entrenamiento. El proceso de validación cruzada es repetido durante  $K$  iteraciones, con cada uno de los posibles

subconjuntos de test. La elección del número de iteraciones depende de la medida del conjunto de datos.

Este método es muy preciso, ya que evaluamos a partir de  $K$  combinaciones de datos de entrenamiento y de test. Tiene una desventaja, y es que es bastante lento desde el punto de vista computacional.

En nuestro caso, debido a que tenemos 40 muestras de datos, lo que vamos a hacer es dividir el conjunto en cinco subconjuntos, cada uno de ellos con ocho carreras distintas. Así el conjunto de entrenamiento contará con cuatro de los subconjuntos y el de test con uno.

De forma que cada iteración será:

- Iteración 1.
  - Conjunto de entrenamiento. Valores comprendidos entre 9 y 40.
  - Conjunto de test. Valores comprendidos entre 1 y 8.
- Iteración 2.
  - Conjunto de entrenamiento. Valores comprendidos entre 1 y 8, y entre 17 y 40.
  - Conjunto de test. Valores comprendidos entre 9 y 16.
- Iteración 3.
  - Conjunto de entrenamiento. Valores comprendidos entre 1 y 16, y entre 25 y 40.
  - Conjunto de test. Valores comprendidos entre 17 y 24.
- Iteración 4.
  - Conjunto de entrenamiento. Valores comprendidos entre 1 y 24, y entre 33 y 40.

- Conjunto de test. Valores comprendidos entre 25 y 32.
- Iteración 5.
  - Conjunto de entrenamiento. Valores comprendidos entre 1 y 32.
  - Conjunto de test. Valores comprendidos entre 33 y 40.

**2. Validación cruzada *Leave-one-out* (LOO).** En este caso lo que vamos a hacer es separar los datos de forma que para cada iteración tengamos sólo una muestra de test y el resto lo usemos como datos de entrenamiento.

En este caso el error es muy bajo, sin embargo es muy costoso a nivel computacional, ya que tiene que realizar un gran número de iteraciones, tantas como muestras tengamos.

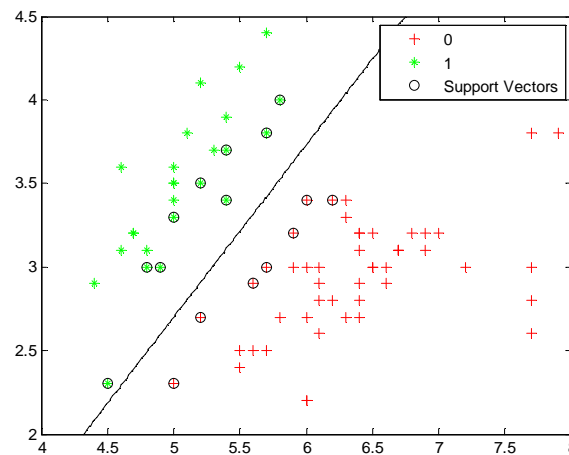
En nuestro caso tendremos 40 iteraciones siendo estas:

- Iteración 1.
  - Conjunto de entrenamiento. Valores comprendidos entre 1 y 39.
  - Conjunto de test. Valor 40.
- Iteración 2.
  - Conjunto de entrenamiento. Valores comprendidos entre 1 y 38 y el 40.
  - Conjunto de test. Valor 39.
- Iteración  $K$ .
  - Conjunto de entrenamiento. Valores comprendidos entre 1 y 40 excepto el valor  $K$ .
  - Conjunto de test. Valor  $K$ .

Una vez que tenemos divididos los datos entre el conjunto de test y el de entrenamiento vamos a realizar una SVM en los conjuntos de entrenamiento.

La SVM [71] es un conjunto de algoritmos de aprendizaje supervisado (técnica para deducir una función a partir de datos de entrenamiento). Dado un conjunto de muestras de entrenamiento podemos etiquetar las clases y entrenar la SVM para construir un modelo que prediga la clase de una nueva muestra. Una SVM es un modelo que representa a los puntos de muestra en el espacio, y se encarga de separar las clases por un espacio lo más amplio posible. Las nuevas muestras (de test) serán clasificadas en una clase o en otra en función de su proximidad a las mismas. Cuanta más separación haya entre clases, mejor será la clasificación realizada

La forma de trabajar de la SVM es construyendo un hiperplano que separe de forma óptima los dos conjuntos distintos de clases. Esta separación óptima se debe a que la SVM busca el hiperplano que tenga la máxima distancia con los puntos que están más cerca de él. De esta forma los puntos del vector que se etiquetan como de una categoría quedan a un lado del hiperplano y los de la otra categoría en el otro lado. Al vector formado por los puntos más cercanos al hiperplano se le llama vector de soporte.



**Figura 2.2.** Ejemplo de hiperplano con función SVM

La construcción del hiperplano hay que realizarla evitando dos fenómenos:

- El subajuste: se basa en obtener de las muestras menos información estadística de la que podríamos sacar. Es decir, realizamos una frontera más sencilla de lo que debería.
- El sobreajuste: ajustamos en exceso los datos, construyendo una frontera excesivamente compleja. De esta forma perdemos la capacidad de generalización, es decir, para los datos que ya tenemos (datos de entrenamiento) funciona muy bien, pero para los nuevos (datos de test) no será válida.

En nuestro desarrollo hemos empleado el modelo de dos etapas descrito en la ecuación (2.13) del artículo expuesto antes. Es decir,

$$\begin{aligned}
 f_{SVM}(x) &= \left( \sum_{i \in SV} \hat{\alpha}_i y_i \exp(-\gamma \|x_i - x\|^2) \right) + \hat{b} \\
 \text{Paso 1: } \{\hat{\alpha}, \hat{b}\} &\leftarrow \max_{\alpha} = \sum_{i=1}^{M_1} \alpha_i \\
 &- \frac{1}{2} \sum_{i,k=1}^{M_1} \alpha_i \alpha_k y_i y_k \exp(-\gamma \|x_i - x_k\|^2) \\
 &s.t. \sum_{i=1}^{M_1} \alpha_i y_i = 0; \quad 0 \leq \alpha_i \leq C \quad \forall i = 1, \dots, M_1 \\
 \text{Paso 2: } p_i^j &= \frac{\beta_1 \exp(f_{SVM}(x_i^j))}{\sum_{i=1}^{m_j} \exp(f_{SVM}(x_i^j)) + \beta_2 q_i^j}, \\
 &i = 1, \dots, M_2; \quad j = 1, \dots, R_2
 \end{aligned}$$

Para realizar la SVM hemos utilizado la función *svmtrain* [70], la cual tiene como parámetros:

- *Boxconstraint* o parámetro *c*. Es el parámetro de suavizado. En nuestro caso el parámetro *c* va a tomar los valores 1, 2.5, 4, 10, 25, 50 y 100. Hemos elegido varios valores, muy distintos entre sí, para ver como varía el resultado

- *KernelCacheLimit*. Valor que especifica el tamaño de memoria caché que vamos a usar para realizar la ejecución de nuestro programa. Cuanto más grande, más rápida será la ejecución, sin embargo, más memoria gastará hasta tal punto de poder interrumpirse la ejecución. Por lo tanto, hay que llegar a un compromiso entre velocidad y consumo de recursos. Nosotros hemos empleado el valor 1000 como tamaño medio de la memoria caché.
- *Kernel\_Function*. En este punto es donde asignamos la función *Kernel* que vamos a usar para realizar el entrenamiento. Hay varios tipos:
  - Lineal. Emplea una función *Kernel* lineal, es decir, se va a realizar un producto escalar.
  - Cuadrática. Función *Kernel* cuadrática.
  - Polinómica. Función *Kernel* polinómica. Por defecto de grado 3, si se quiere otro grado hay que especificarlo.
  - RBF. Función *Kernel* Gaussiana de Base Radial, como la que hemos descrito en el artículo en el punto 3.1. Como hemos dicho que vamos a seguir el modelo, nuestra función *Kernel* va a pertenecer a esta clase.
- *Rbf\_Sigma*. En las funciones *Kernel* RBF vamos a necesitar un factor de escala. Aquí es donde pondremos dicho factor.

Para entrenar a la máquina, vamos a usar varios valores de sigma. Estos son 0,4145; 0,829; 1,658; 3,316; 6,632; 13,26 y 15. Al igual que en el caso del parámetro  $c$ , aquí también hemos escogido valores muy dispares.

- *Method*. Se refiere al método que vamos a usar para encontrar el hiperplano de separación. Hay varias opciones:
  - QP. Se refiere a programación cuadrática.
  - SMO. Optimización mínima secuencial, que es un algoritmo mejorado de SVM que trabaja con problemas cuadráticos grandes, los cuales los divide en



problemas más pequeños. De esta forma el tiempo de cómputo es menor y minimiza el error de generalización. Nosotros vamos a usar este método, ya que así se resolverá de forma más rápida.

- LS. Es el método de mínimos cuadrados.

Cuando usamos el método SMO, le tenemos que introducir una serie de parámetros, estos van a ser:

- *Display*. Especifica el nivel de información acerca de las iteraciones de optimización que muestran como se ejecuta el programa. Tiene como opciones:
  - *Off*. No reporta ningún tipo de información.
  - *Iter*. Realiza un informe cada quinientas iteraciones.
  - *Final*. Sólo se informa cuando se termina de ejecutar el programa.
- *MaxIter*. Es un número entero que especifica el número de iteraciones que va a tener nuestro programa. En este caso también hay que llegar a un compromiso, ya que si el número es demasiado pequeño y se supera este límite antes de que el algoritmo converja, entonces el algoritmo se detiene y devuelve un error. Sin embargo, si el número de iteraciones es muy grande el programa tardará mucho en realizar la ejecución. En nuestro caso este número será 200.000, ya que es un programa muy grande, y por lo tanto, con este número nos aseguramos de que no vamos a tener ningún error de ejecución.

Una vez que hemos efectuado la función SVM, tenemos que realizar una función *Kernel* [71]. Esto lo vamos a tener que usar debido a las limitaciones computacionales de las máquinas de aprendizaje lineal. La representación por medio de funciones *Kernel* se basa en proyectar la información a un espacio de características de mayor dimensión, el cuál aumenta la capacidad computacional de las máquinas de aprendizaje lineal. Es decir, mapearemos el espacio de entradas a un nuevo espacio de mayor dimensionalidad.

Teniendo estos datos ya podemos pasar a la segunda etapa del modelo que hemos estado siguiendo. Es decir, calcular, a través de lo hallado anteriormente, la probabilidad que tiene cada caballo de ganar dentro de cada carrera.

$$\text{Paso 2: } p_i^j = \frac{\beta_1 \exp(f_{SVM}(x_i^j))}{\sum_{i=1}^{m_j} \exp(f_{SVM}(x_i^j)) + \beta_2 q_i^j},$$

$$i = 1, \dots, M_2; \quad j = 1, \dots, R_2$$

Para calcular el valor de  $\beta_1$  y  $\beta_2$  vamos a tomar en consideración que  $\beta_2 = (1 - \beta_1)$ . De esta forma, nos vamos a encontrar con distintas situaciones:

- $\beta_1 = 1$  y  $\beta_2 = 0$ . En este caso, sólo vamos a tener en cuenta los resultados calculados a través de la SVM.
- $\beta_2 = 1$  y  $\beta_1 = 0$ . Aquí vamos a utilizar únicamente las predicciones realizadas por los expertos.
- $0 < \beta_1 < 1$ . En este caso vamos a calcular las probabilidades con ambas predicciones. Dando más peso a una o a otra dependiendo de cómo de grande sea  $\beta_1$ .

#### 2.2.4. Odds y Factor Q

El cálculo de las *odds* a partir de las probabilidades de cada caballo se hace con:

$$O_i = \frac{QW}{P_i} - 1 \quad (2.14)$$

En la que:

- $O_i$  son las *odds* de la casa de apuestas.
- $P_i$  es la probabilidad del caballo  $i$  de ganar la carrera según el mercado.
- $Q$  es el porcentaje en tanto por uno de lo que se reparte. Es decir, hay un tanto por ciento que se queda la casa de apuestas, y la  $Q$  es el restante, que va a ser repartido entre todos los acertantes de las apuestas.

- $W = 1$

Si lo comparamos con la misma fórmula en el caso del fútbol,

$$O_i = \frac{QW}{P_i}$$

Nos encontramos con que en este caso no tenemos un factor 1 restando, esto es así, porque en el caso de las carreras de caballos las casas de apuestas cuando haces una combinación ganadora, además de darte lo correspondiente por haber ganado, te devuelven lo apostado. Por ejemplo, en el caso de que apuestes 0,30 céntimos por el caballo que queda primero, si la ganancia es de 1€, el total que ganarás será de 1,30€. En el caso del fútbol sólo ganarías los 0,30 céntimos (no te devuelven lo apostado).

Por otro lado también sabemos que:

$$\sum_{i=1}^N P_i = 1 \quad (2.15)$$

Es decir, la suma de las probabilidades de ganar de cada caballo en una carrera tiene que ser 1.

Por lo tanto, tenemos un sistema de  $N+1$  ecuaciones (2.14 y 2.15) con  $N+1$  incógnitas ( $Q$  y  $P_N$ ).

Si despejamos (2.14) tenemos que:

$$\begin{aligned} (O_1 + 1) \cdot p_1 - Q &= 0 \\ (O_2 + 1) \cdot p_2 - Q &= 0 \\ &\vdots \\ (O_N + 1) \cdot p_N - Q &= 0 \end{aligned}$$

Por lo tanto, vamos a resolver estas ecuaciones a través de un sistema matricial de la forma siguiente:

$$\begin{pmatrix} O_1+1 & 0 & \dots & 0 & -1 \\ 0 & O_2+1 & 0 & \vdots & -1 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & 0 & O_N+1 & \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix} \bullet \begin{pmatrix} p_1 \\ \vdots \\ p_N \\ Q \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

↓

$$\begin{pmatrix} O_1+1 & 0 & \dots & 0 & -1 \\ 0 & O_2+1 & 0 & \vdots & -1 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & 0 & O_N+1 & \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} p_1 \\ \vdots \\ p_N \\ Q \end{pmatrix}$$

Es decir, para calcular el vector de  $P$  y la  $Q$  tenemos que hacer la inversa de la matriz de *odds*.

De esta forma ya tendríamos todos los datos necesarios para calcular nuestra probabilidad de un caballo de ganar en una determinada carrera.

Vamos a poner un ejemplo de lo mostrado anteriormente, para que se entienda mejor.

Tenemos una carrera con diez caballos. Cada caballo tiene las siguientes *odds* según *Formstar*:

	Caballo 1	Caballo 2	Caballo 3	Caballo 4	Caballo 5	Caballo 6	Caballo 7	Caballo 8	Caballo 9	Caballo 10
Odds	3,5	3,8	9	6,5	5,5	13	21	16	34	67

**Tabla 2.11.** Ejemplo con las *odds* de los caballos de una carrera

Si sustituímos en la última matriz del desarrollo:

$$\begin{pmatrix} 3,5+1 & 0 & \dots & 0 & -1 \\ 0 & 3,8+1 & 0 & \vdots & -1 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & 0 & 67+1 & \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} p_1 \\ \vdots \\ p_{10} \\ Q \end{pmatrix}$$

$$\Downarrow$$

$$\begin{pmatrix} 4,5 & 0 & \dots & 0 & -1 \\ 0 & 4,8 & 0 & \vdots & -1 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & 0 & 68 & \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} p_1 \\ \vdots \\ p_{10} \\ Q \end{pmatrix}$$

Tendremos que:

- Caballo 1.  $p_1 = 0,2679$
- Caballo 2.  $p_2 = 0,1705$
- Caballo 3.  $p_3 = 0,1443$
- Caballo 4.  $p_4 = 0,0938$
- Caballo 5.  $p_5 = 0,0938$
- Caballo 6.  $p_6 = 0,0853$
- Caballo 7.  $p_7 = 0,0721$
- Caballo 8.  $p_8 = 0,0447$
- Caballo 9.  $p_9 = 0,0184$
- Caballo 10.  $p_{10} = 0,0093$
- Valor Q:  $Q = 0,9378$ . Es decir, la casa de apuestas *Formstar* se queda con un 6,22% de la cantidad que apuestan los usuarios en la misma.

	Caballo 1	Caballo 2	Caballo 3	Caballo 4	Caballo 5	Caballo 6	Caballo 7	Caballo 8	Caballo 9	Caballo 10
p	0,2679	0,1705	0,1443	0,0938	0,0938	0,0853	0,0721	0,0447	0,0184	0,0093

**Tabla 2.12.** Ejemplo con las probabilidades de la página web para cada caballo en una carrera

Una vez que ya tenemos el valor  $q_i^j$  podemos sustituir en la ecuación (2.13) y, de esta forma, calcular nuestras probabilidades de que un caballo gane en una carrera.

Para el ejemplo anterior, las probabilidades finales de cada caballo serán:

- Caballo 1.  $p_1 = 0,0721$
- Caballo 2.  $p_2 = 0,0519$
- Caballo 3.  $p_3 = 0,0648$
- Caballo 4.  $p_4 = 0,0600$
- Caballo 5.  $p_5 = 0,0440$
- Caballo 6.  $p_6 = 0,0884$
- Caballo 7.  $p_7 = 0,0805$
- Caballo 8.  $p_8 = 0,0773$
- Caballo 9.  $p_9 = 0,1833$
- Caballo 10.  $p_{10} = 0,2777$

	Caballo 1	Caballo 2	Caballo 3	Caballo 4	Caballo 5	Caballo 6	Caballo 7	Caballo 8	Caballo 9	Caballo 10
p	0,0721	0,0519	0,0648	0,0600	0,0440	0,0884	0,0805	0,0773	0,1833	0,2777

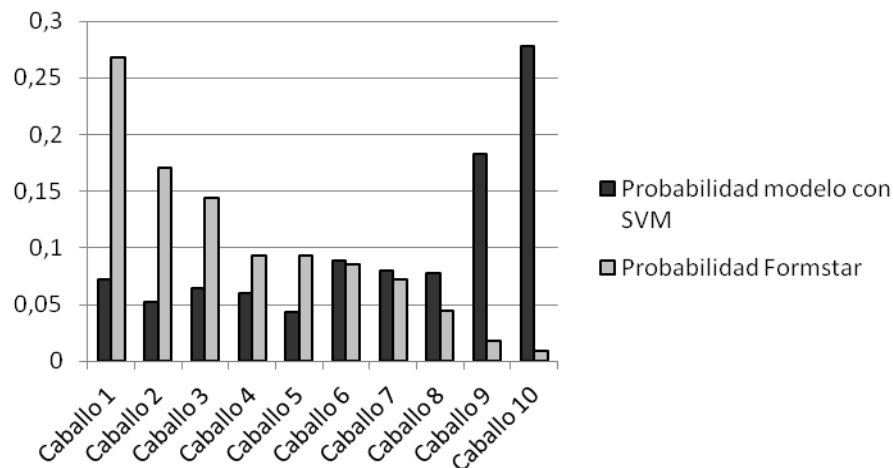
**Tabla 2.13.** Ejemplo con nuestras probabilidades, calculadas a través de las *odds*, para cada caballo en una carrera

Para tener más ganancias, uno de los objetivos era que la probabilidad de ganar de cada caballo calculada a través de la SVM, fuera distinta a la que han calculado en la casa de apuestas en la que nos estamos basando. De esta forma, si encontramos un caballo que tenga, según nosotros, más posibilidades de las que se están contemplando por el resto, nuestros beneficios serán mayores.

En este ejemplo nos encontramos con:

	Caballo 1	Caballo 2	Caballo 3	Caballo 4	Caballo 5	Caballo 6	Caballo 7	Caballo 8	Caballo 9	Caballo 10
Probabilidad modelo con SVM	0,0721	0,0519	0,0648	0,06	0,044	0,0884	0,0805	0,0773	0,1833	0,2777
Probabilidad Formstar	0,2679	0,1705	0,1443	0,0938	0,0938	0,0853	0,0721	0,0447	0,0184	0,0093

**Tabla 2.14.** Comparación de nuestras probabilidades con las de *Formstar*, para cada caballo, en una carrera



**Figura 2.3.** Comparación de nuestras probabilidades con las de *Formstar*, para cada caballo, en una carrera

Como vemos, nos sería más rentable apostar a los caballos 6, 7, 8, 9 y 10. Sobre todo, si apostamos por los caballos 9 y 10, y algunos de esos dos resulta ser el ganador, nuestro beneficio sería muy grande, ya que estamos apostando por un caballo que en realidad, para el resto de apostantes no tenía casi ninguna posibilidad de ganar, y por lo tanto, nuestra cuota se verá multiplicada por un número mayor.

### 2.2.5. Criterio de Kelly

El Criterio de Kelly [61] es una herramienta muy útil en el mundo de las apuestas, ya que nos dice cuanto debemos apostar para obtener beneficios.

La fórmula del Criterio de Kelly fue desarrollada en 1956 por John Kelly [72], un científico norteamericano cuyo objetivo era investigar sobre la mejor opción para gestionar una cartera monetaria. Es utilizado en la bolsa para realizar las carteras de valores en función de las ganancias esperadas de cada título y las probabilidades estimadas para un período de tiempo concreto. También es usado por jugadores expertos de Poker y BlackJack, así como por jugadores de casas de apuestas y en general, de toda clase de juegos.

Está diseñado para maximizar el crecimiento de nuestro dinero o *bankroll* disponible a largo plazo, estimando las cantidades óptimas a jugar en cada apuesta o *stake*. Para que este método sea rentable, debemos ser capaces de realizar unas estimaciones de las probabilidades mejores que las de las casas de apuestas. Cuanto más distintas sean nuestras probabilidades de las de las casas de apuestas, mayor será nuestra ganancia.

Hay apostantes que consideran este criterio demasiado arriesgado, puesto que su efectividad depende de que nuestras predicciones sean mejores que las de las casas de apuestas, y eso es muy complicado. Por ello se usan variantes llamadas Fraccionario de Kelly en las que se apuesta a la mitad o a un cuarto de lo recomendado, así se minimiza un poco el riesgo de perder, pero los beneficios que se obtienen a largo plazo son menores.

Algo muy importante de la utilización del Criterio de Kelly es que utilizando esta estrategia nunca llegaremos a la bancarrota, ya que la cantidad a apostar se determina mediante un porcentaje del *bankroll*.

Existen otros sistemas de gestión de banca, pero ninguno revolucionó tanto el mundo de las apuestas como el Criterio de Kelly.



Para utilizar el Criterio de Kelly lo primero que tenemos que hacer es calcular nuestras probabilidades (en nuestro caso aplicando una SVM). También necesitamos los retornos de la casa de apuestas ( $\alpha$ ) donde tenemos la intención de realizar el *stake*.

Este criterio se puede resumir en las siguientes etapas [73]:

1. Tomando como conocidas los retornos de las casas de apuestas y la probabilidad calculada por nosotros anteriormente, tenemos que realizar la multiplicación entre ambas, para cada uno de los casos, y ordenarlos de mayor a menor. De forma que

$$p(s) \cdot \alpha_s \geq p(s+1) \cdot \alpha_{s+1}$$

Siendo  $s$  un determinado evento.

2. Cuando tenemos ordenados los índices, debemos calcular  $b$ . Siendo este el menor valor mayor que 0 resultado de:

$$F_t = \frac{1 - p_t}{1 - \sigma_t} \quad t = \text{Evento1}, \text{Evento2} \dots \text{Eventot}$$

Donde:

$$p_t = \sum_1^t p(s), \quad \sigma_t = \sum_1^t \frac{1}{\alpha_s} \quad y \quad F_0 = 1$$

3. Por último calculamos el vector de los *stakes*, o lo que es lo mismo, el vector que devuelve la cantidad de dinero que vamos a apostar a cada evento.

$$a(s) = p(s) - \frac{b}{\alpha_s}$$

Para realizar una mayor explicación del mismo, vamos a ilustrarlo con el ejemplo de una carrera que cuenta con diez caballos.

Partimos con los siguientes valores:

- $\alpha \rightarrow$  Cuotas de cada caballo en una carrera, dadas por *Formstar*.

- $p \rightarrow$  Probabilidades obtenidas mediante la SVM

	Caballo 1	Caballo 2	Caballo 3	Caballo 4	Caballo 5	Caballo 6	Caballo 7	Caballo 8	Caballo 9	Caballo 10
p	0,0721	0,0519	0,0648	0,0600	0,0440	0,0884	0,0805	0,0773	0,1833	0,2777
$\alpha$	3,5	3,8	9	6,5	5,5	13	21	16	34	67

**Tabla 2.15.** Valores de las probabilidades obtenidas con SVM y de las cuotas de cada caballo en una carrera para Criterio de Kelly

Una vez que tenemos estos dos datos, hay que realizar la multiplicación de ambos y ordenarlos de mayor a menor.

- Caballo 1:  $0,00721 \cdot 3,5 = 0,2525$
- Caballo 2:  $0,0519 \cdot 3,8 = 0,1971$
- Caballo 3:  $0,0648 \cdot 9 = 0,5824$
- Caballo 4:  $0,0600 \cdot 6,5 = 0,3901$
- Caballo 5:  $0,0440 \cdot 5,5 = 0,2422$
- Caballo 6:  $0,0884 \cdot 13 = 1,1488$
- Caballo 7:  $0,0805 \cdot 21 = 1,6910$
- Caballo 8:  $0,0773 \cdot 16 = 1,2366$
- Caballo 9:  $0,1833 \cdot 34 = 6,2331$
- Caballo 10:  $0,2777 \cdot 67 = 18,6068$

De forma ordenada quedarían: Caballo 10 > Caballo 9 > Caballo 7 > Caballo 8 > Caballo 6 > Caballo 3 > Caballo 4 > Caballo 1 > Caballo 5 > Caballo 2

Luego tenemos que aplicar la siguiente fórmula

$$F_t = \frac{1 - p_t}{1 - \sigma_t} \quad t = \text{Caballo1}, \text{Caballo2} \dots \text{Caballot}$$

El orden de t no lo conocemos con anticipación, en nuestro caso depende del número de caballos que corran la carrera a estudiar.

- Caballo 10

$$F(\text{Caballo10}) = \frac{1 - 0,2777}{1 - \frac{1}{67}} = 0,7332$$

- Caballo 9

$$F(\text{Caballo9}) = \frac{1 - 0,2777 - 0,1833}{1 - \frac{1}{67 + 34}} = 0,5640$$

- Caballo 7

$$F(\text{Caballo7}) = \frac{1 - 0,2777 - 0,1833 - 0,0805}{1 - \frac{1}{67 + 34 + 21}} = 0,5049$$

- Caballo 8

$$F(\text{Caballo8}) = \frac{1 - 0,2777 - 0,1833 - 0,0805 - 0,0773}{1 - \frac{1}{67 + 34 + 21 + 16}} = 0,4508$$

- Caballo 6

$$F(\text{Caballo6}) = \frac{1 - 0,2777 - 0,1833 - 0,0805 - 0,0773 - 0,0804}{1 - \frac{1}{67 + 34 + 21 + 16 + 13}} = 0,3809$$

- Caballo 3

$$F(\text{Caballo 3}) = \frac{1 - 0,2777 - 0,1833 - 0,0805 - 0,0773 - 0,0804 - 0,0648}{1 - \frac{1}{67 + 34 + 21 + 16 + 13 + 8}}$$

$$F(\text{Caballo 3}) = 0,3469$$

- Caballo 4

$$F(\text{Caballo 4}) = \frac{1 - 0,2777 - 0,1833 - 0,0805 - 0,0773 - 0,0804 - 0,0648 - 0,0600}{1 - \frac{1}{67 + 34 + 21 + 16 + 13 + 8 + 6,5}}$$

$$F(\text{Caballo 4}) = 0,3337$$

- Caballo 1

$$F(\text{Caballo 1}) = \frac{1 - 0,2777 - 0,1833 - 0,0805 - 0,0773 - 0,0804 - 0,0648 - 0,0600 - 0,0721}{1 - \frac{1}{67 + 34 + 21 + 16 + 13 + 8 + 6,5 + 5,5}}$$

$$F(\text{Caballo 1}) = 0,4400$$

- Caballo 5

$$F(\text{Caballo 2}) = \frac{1 - 0,2777 - 0,1833 - 0,0805 - 0,0773 - 0,0804 - 0,0648 - 0,0600 - 0,0721 - 0,0440}{1 - \frac{1}{67 + 34 + 21 + 16 + 13 + 8 + 6,5 + 5,5 + 3,8}}$$

$$F(\text{Caballo 2}) = 1,4357$$

Calculamos el valor de  $b$ , este es 0,3337, además se encuentra en la posición  $F_7$ .

Por último tenemos que obtener los *stakes*. Vamos a apostar hasta el caballo que en el vector ordenado estaba en la posición 7, por lo tanto, apostaremos por el Caballo 10, Caballo 9, Caballo 7, Caballo 8, Caballo 6, Caballo 3 y Caballo 4. Así nos queda que:

- Caballo 1:  $a_1 = 0$

- Caballo 2:  $a_2 = 0$

- Caballo 3:

$$a_3 = p_3 - \frac{b}{\alpha_3} = 0,0647 - \frac{0,3337}{9} = 0,0276$$

- Caballo 4:

$$a_4 = p_4 - \frac{b}{\alpha_4} = 0,0440 - \frac{0,3337}{6,5} = 0,0087$$

- Caballo 5:  $a_5 = 0$

- Caballo 6:

$$a_6 = p_6 - \frac{b}{\alpha_6} = 0,0884 - \frac{0,3337}{13} = 0,0627$$

- Caballo 7:

$$a_7 = p_7 - \frac{b}{\alpha_7} = 0,0805 - \frac{0,3337}{21} = 0,0646$$

- Caballo 8:

$$a_8 = p_8 - \frac{b}{\alpha_8} = 0,0773 - \frac{0,3337}{16} = 0,0564$$

- Caballo 9:

$$a_9 = p_9 - \frac{b}{\alpha_9} = 0,1833 - \frac{0,3337}{34} = 0,1735$$

- Caballo 10:

$$a_{10} = p_{10} - \frac{b}{\alpha_{10}} = 0,2777 - \frac{0,3337}{67} = 0,2727$$

De esta forma nuestro vector  $a(s)$  quedaría:

$$a(s) = [0 \quad 0 \quad 0,0276 \quad 0,0087 \quad 0 \quad 0,0627 \quad 0,0646 \quad 0,0564 \quad 0,1735 \quad 0,2727]$$

El ganador de esta carrera fue el caballo número uno. Por ese caballo no realizamos ninguna apuesta, por lo tanto, en esta carrera habríamos perdido lo apostado, que son 0,6663€.

# CAPÍTULO 3

## Resultados

En este capítulo vamos a mostrar los resultados que hemos obtenido desarrollando el modelo descrito en el capítulo anterior.

Los datos en los que nos hemos basado han sido carreras realizadas entre Febrero y Junio de 2011. Además, como hemos mencionado antes, todos los datos han sido sacados de la página de *Formstar* que cuenta con información de carreras de Sudáfrica.

El estudio se ha basado en las carreras E Division de distancia 1400 metros. Hemos analizado el método haciendo una distinción por distancia para que los resultados obtenidos fueran más acertados.

Para calcular las ganancias tenemos un *bankroll* de 1€ por carrera, independientemente del número de carreras que tengamos.

Además, el análisis lo vamos a hacer con dos tipos de validaciones cruzadas:

- Validación cruzada de  $K$  iteraciones.
- Validación cruzada *Leave-One-Out*.

## 3.1. Elección del valor de $\beta$

Como hemos visto antes, la función SVM va a variar en función del valor que tome  $\beta$ .

Por lo tanto, en esta sección trataremos de elegir el valor óptimo de  $\beta$ .

### 3.1.1. Probabilidad de ganar de un caballo en función del valor de $\beta$

Vamos a mostrar cómo varía la probabilidad que tiene cada caballo de ganar, en una carrera, en función de la elección que se haga de  $\beta$ . Para ello vamos a mostrar una carrera que contiene 15 caballos. El estudio lo vamos a hacer desde el punto de vista de CV y, a continuación, con LOO.

#### 3.1.1.1. Estudio CV

En primer lugar estudiaremos como influye  $\beta$  en el caso de que estemos usando validación cruzada de  $K$  iteraciones.

Vamos a ver cómo varía la probabilidad de ganar de cada caballo, según vamos variando el valor de  $\beta$ . Una vez que tenemos todas las probabilidades, y sabiendo cuál ha sido el ganador de la carrera, podremos decidir cuál es el valor de  $\beta$  que nos habría llevado a realizar una mayor apuesta por ese caballo.

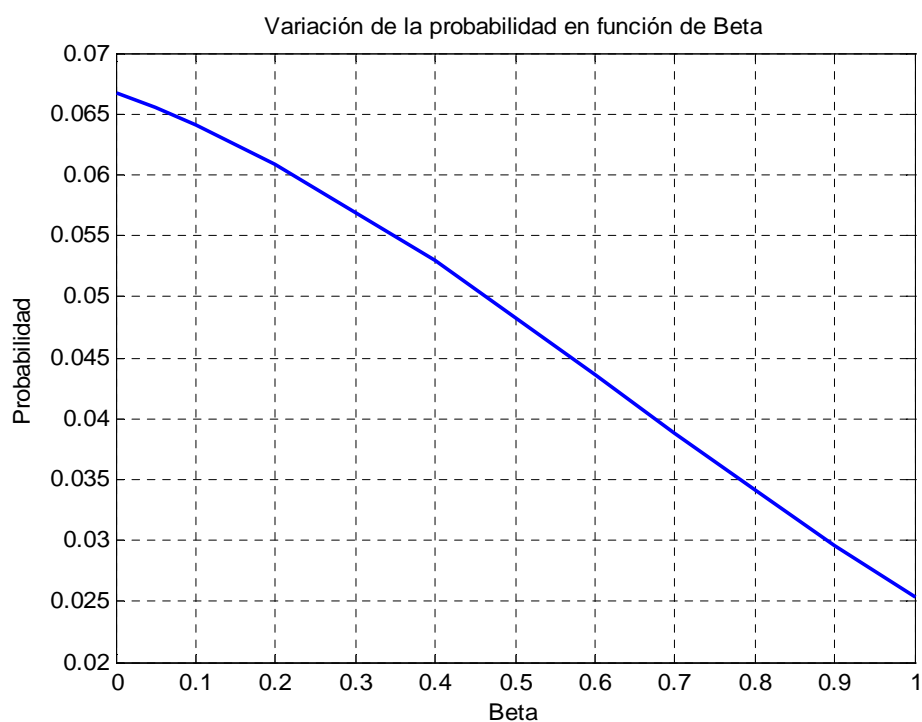


	Beta=0	Beta=0,05	Beta=0,1	Beta=0,2	Beta=0,4	Beta=0,6	Beta=0,7	Beta=0,8	Beta=0,9	Beta=1
Caballo 1	0,0667	0,0655	0,0641	0,0609	0,0529	0,0436	0,0388	0,0341	0,0296	0,0254
Caballo 2	0,0667	0,0633	0,0599	0,0532	0,0403	0,0290	0,0241	0,0198	0,0161	0,0129
Caballo 3	<b>0,0667</b>	0,0634	0,0601	0,0536	0,0410	0,0297	0,0248	0,0205	0,0167	0,0134
Caballo 4	0,0667	0,0689	0,0711	0,0748	0,0798	0,0809	0,0798	0,0777	0,0747	0,0710
Caballo 5	0,0667	0,0639	0,0611	0,0553	0,0436	0,0326	0,0277	0,0232	0,0192	0,0156
Caballo 6	0,0667	0,0616	0,0567	0,0477	0,0324	0,0209	0,0165	0,0128	0,0098	0,0075
Caballo 7	0,0667	0,0614	0,0564	0,0471	0,0317	0,0202	0,0158	0,0122	0,0093	0,0070
Caballo 8	0,0667	0,0666	0,0663	0,0651	0,0604	0,0532	0,0490	0,0445	0,0399	0,0354
Caballo 9	0,0667	0,0699	0,0731	0,0792	0,0896	0,0960	0,0975	0,0977	0,0967	0,0945
Caballo 10	0,0667	0,0653	0,0638	0,0603	0,0519	0,0423	0,0375	0,0328	0,0283	0,0241
Caballo 11	0,0667	0,0751	0,0844	0,1055	0,1589	0,2270	0,2660	0,3076	0,3513	0,3963
Caballo 12	0,0667	0,0688	0,0709	0,0744	0,0790	0,0796	0,0784	0,0761	0,0730	0,0692
Caballo 13	0,0667	0,0711	0,0755	0,0845	0,1018	0,1165	0,1221	0,1264	0,1291	0,1303
Caballo 14	0,0667	0,0668	0,0668	0,0661	0,0624	0,0558	0,0518	0,0474	0,0429	0,0383
Caballo 15	0,0667	0,0683	0,0698	0,0721	0,0742	0,0725	0,0702	0,0672	0,0634	0,0591

**Tabla 3.1.** Ejemplo de las probabilidades de ganar de cada caballo por cada  $\beta$  con validación cruzada

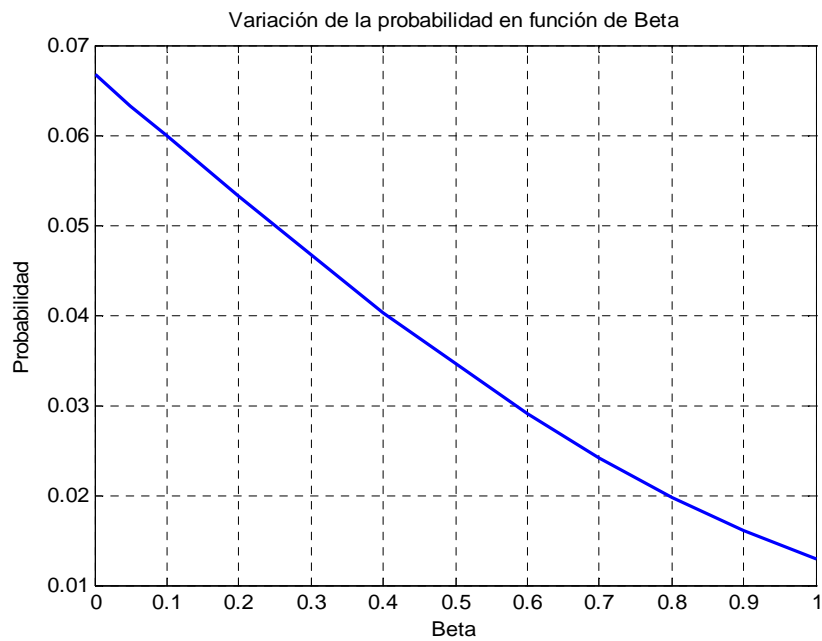
Para mayor facilidad, vamos a mostrarlo gráficamente:

- Caballo 1



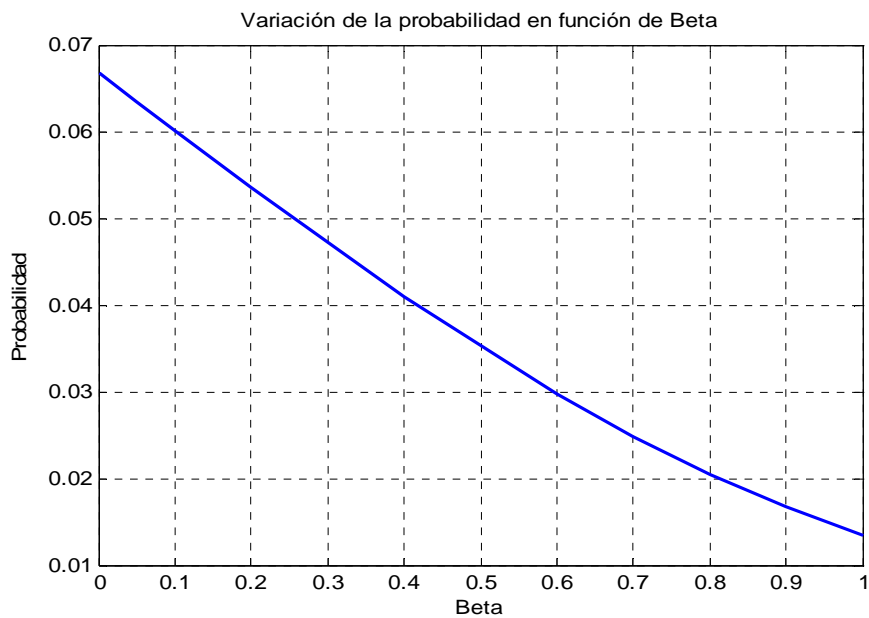
**Figura 3.1.** Variación de la probabilidad de ganar del caballo 1 en función de  $\beta$  con validación cruzada

- Caballo 2



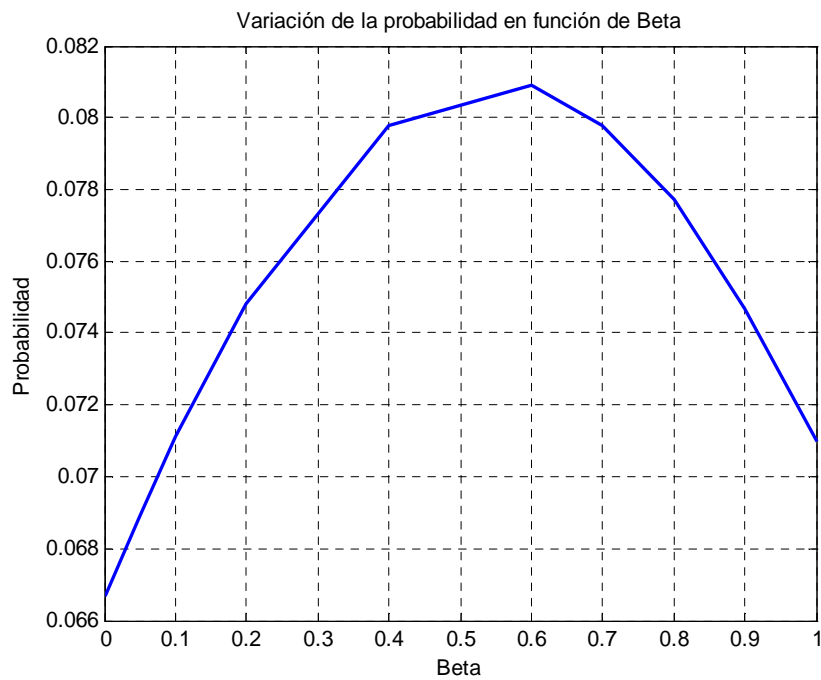
**Figura 3.2.** Variación de la probabilidad de ganar del caballo 2 en función de  $\beta$  con validación cruzada

- Caballo 3



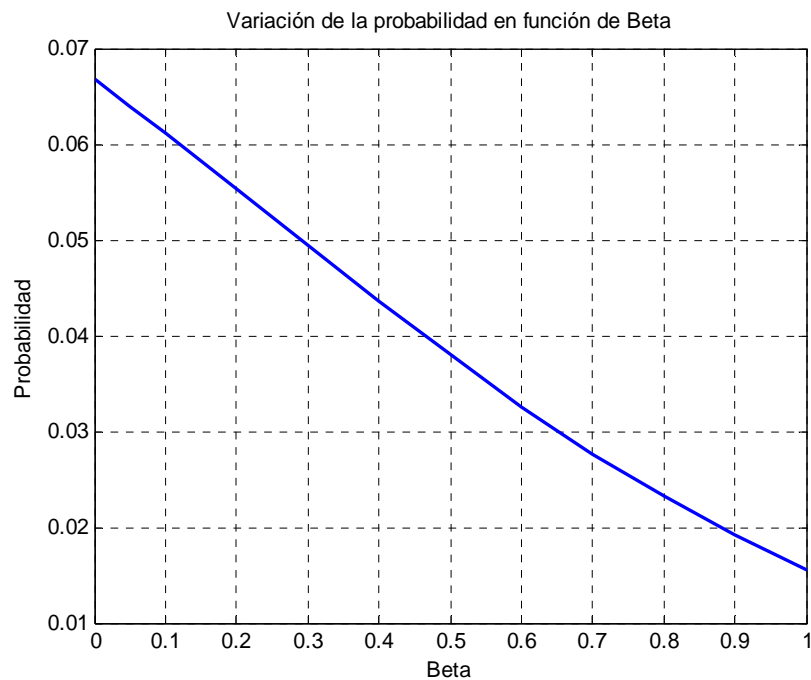
**Figura 3.3.** Variación de la probabilidad de ganar del caballo 3 en función de  $\beta$  con validación cruzada

- Caballo 4



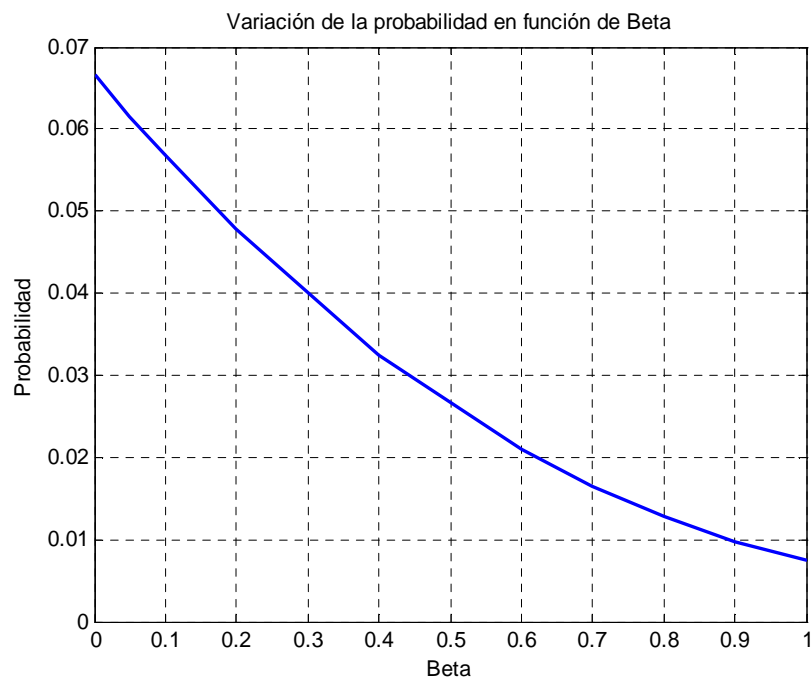
**Figura 3.4.** Variación de la probabilidad de ganar del caballo 4 en función de  $\beta$  con validación cruzada

- Caballo 5



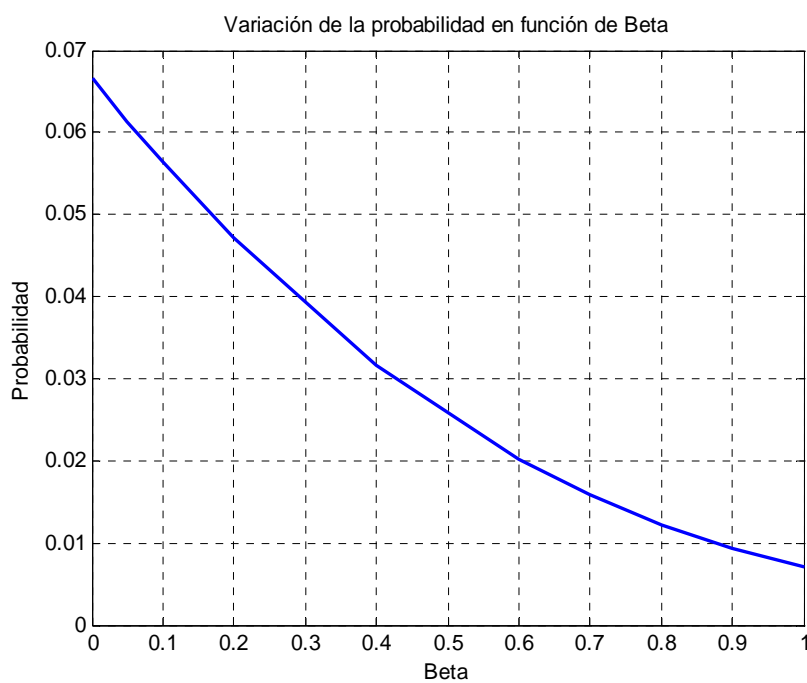
**Figura 3.5.** Variación de la probabilidad de ganar del caballo 5 en función de  $\beta$  con validación cruzada

- Caballo 6



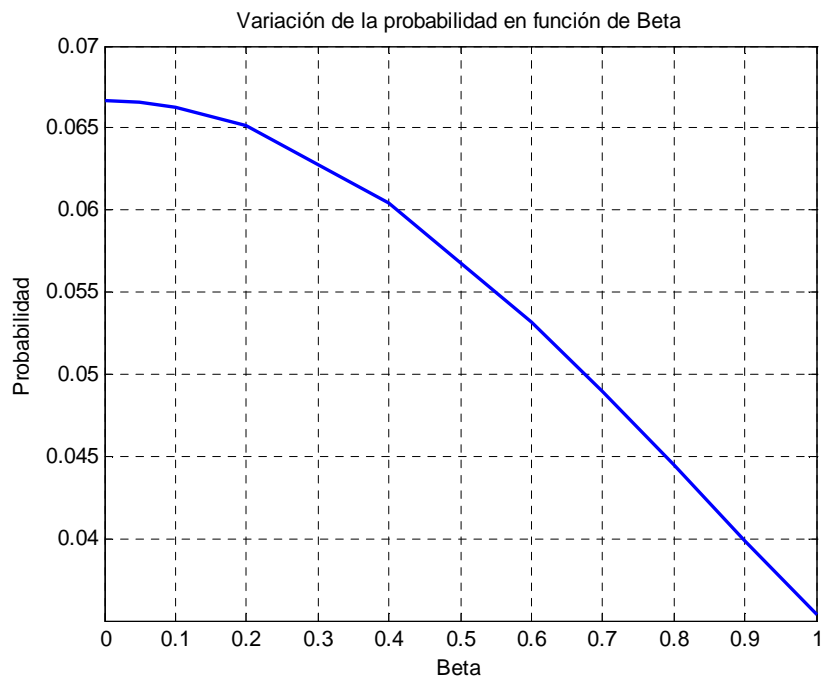
**Figura 3.6.** Variación de la probabilidad de ganar del caballo 6 en función de  $\beta$  con validación cruzada

- Caballo 7



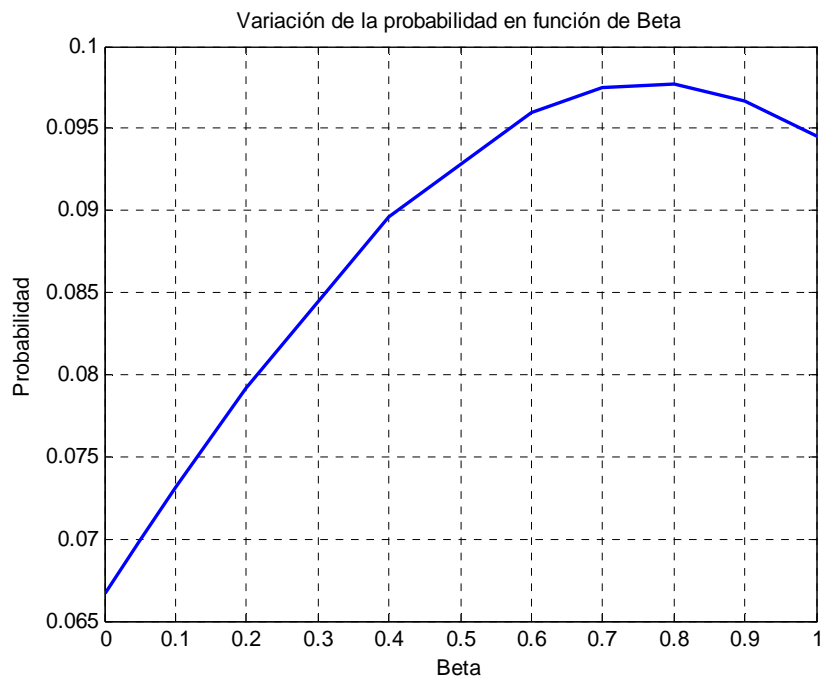
**Figura 3.7.** Variación de la probabilidad de ganar del caballo 7 en función de  $\beta$  con validación cruzada

- Caballo 8



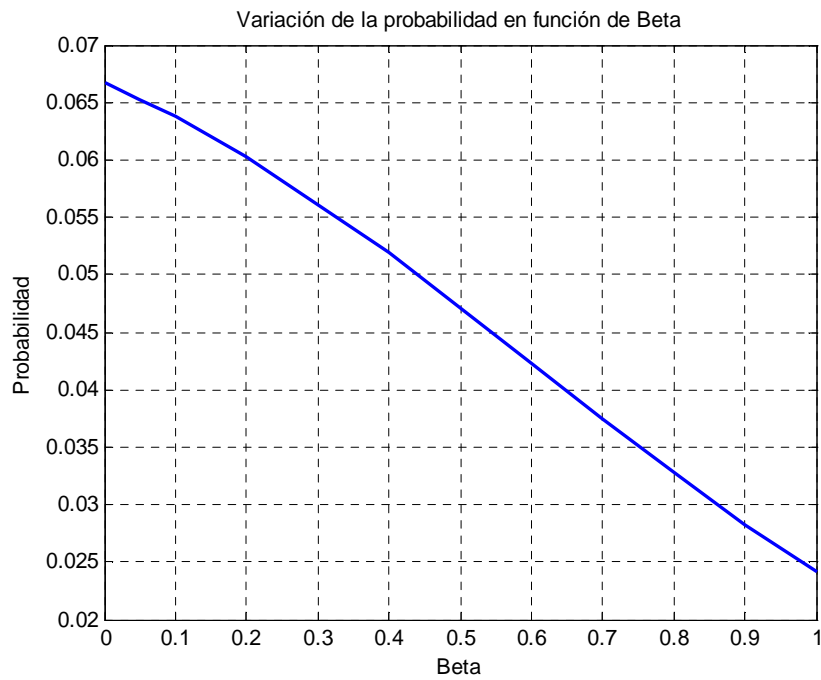
**Figura 3.8.** Variación de la probabilidad de ganar del caballo 8 en función de  $\beta$  con validación cruzada

- Caballo 9



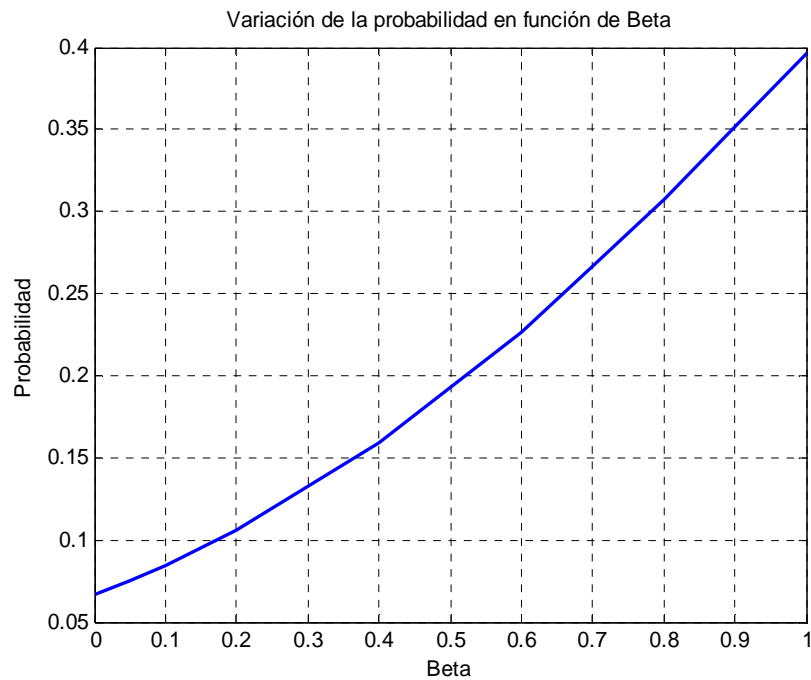
**Figura 3.9.** Variación de la probabilidad de ganar del caballo 9 en función de  $\beta$  con validación cruzada

- Caballo 10



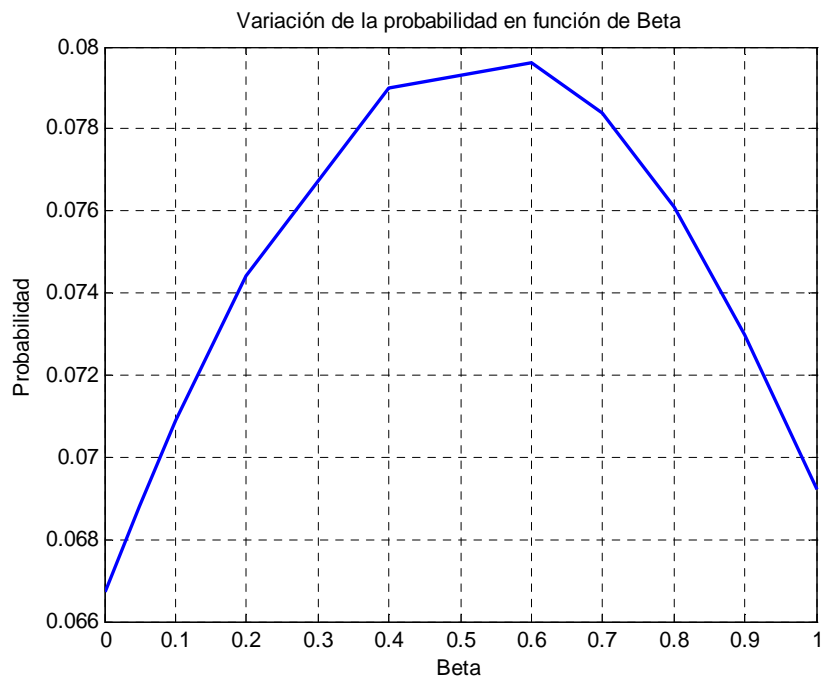
**Figura 3.10.** Variación de la probabilidad de ganar del caballo 10 en función de  $\beta$  con validación cruzada

- Caballo 11



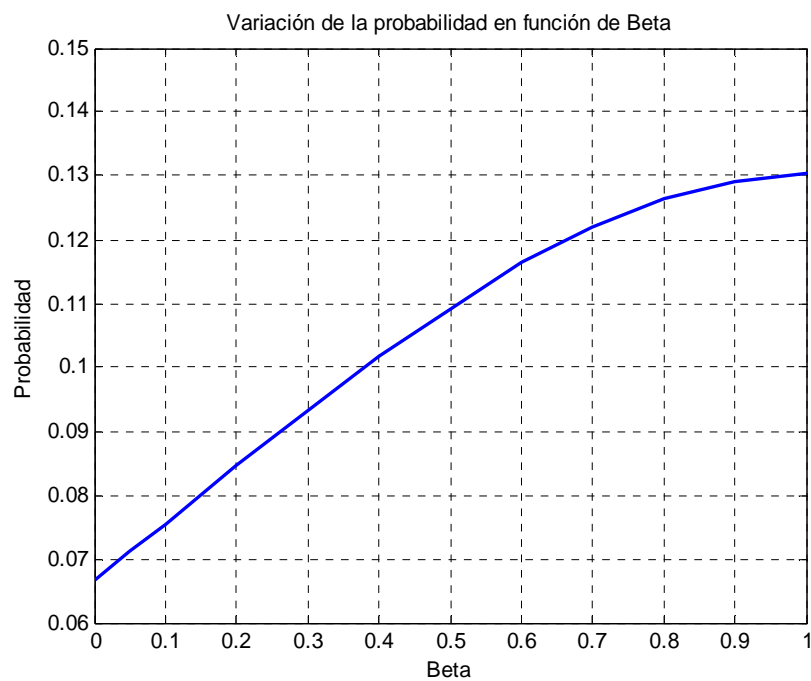
**Figura 3.11.** Variación de la probabilidad de ganar del caballo 11 en función de  $\beta$  con validación cruzada

- Caballo 12



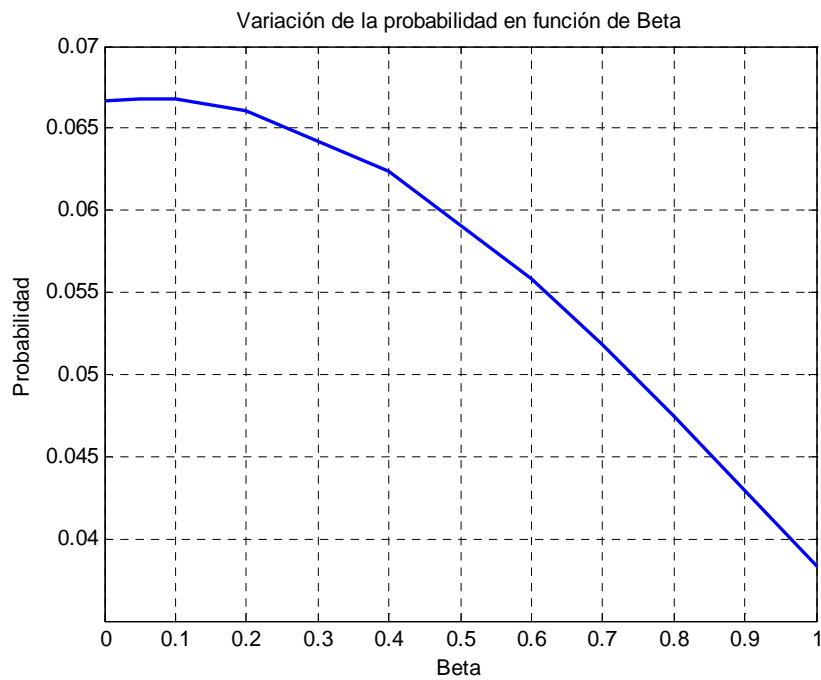
**Figura 3.12.** Variación de la probabilidad de ganar del caballo 12 en función de  $\beta$  con validación cruzada

- Caballo 13



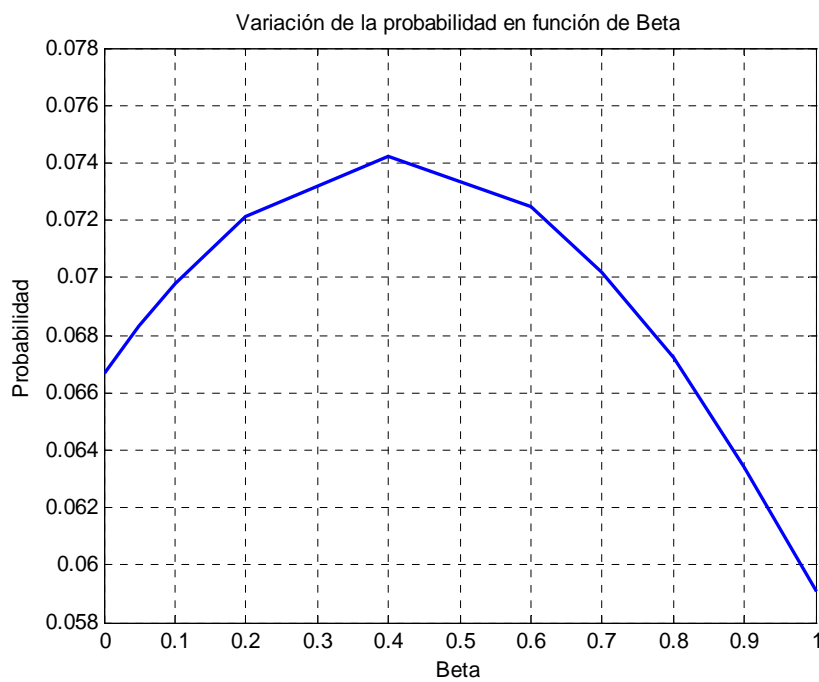
**Figura 3.13.** Variación de la probabilidad de ganar del caballo 13 en función de  $\beta$  con validación cruzada

- Caballo 14



**Figura 3.14.** Variación de la probabilidad de ganar del caballo 14 en función de  $\beta$  con validación cruzada

- Caballo 15



**Figura 3.15.** Variación de la probabilidad de ganar del caballo 15 en función de  $\beta$  con validación cruzada



Como vemos, las anteriores gráficas no siguen un patrón. Tenemos el ejemplo del caballo 1, caballo 2, caballo 3, caballo 5, caballo 6, caballo 7, caballo 8, caballo 10 y caballo 14 que tienen su máximo con  $\beta = 0$  y desde ahí van bajando hasta llegar al mínimo en  $\beta = 1$ . Luego los caballos 4, 12 y 15 empiezan en un valor alto con  $\beta = 0$ , pero, en lugar de ir descendiendo, suben teniendo su máximo en  $\beta = 0,6$  los dos primeros y  $\beta = 0,4$  el último, para, a continuación, ir hasta su mínimo en  $\beta = 1$ . Y por último están los caballos 9, 11 y 13 que en  $\beta = 0$  tienen su valor más pequeño y van ascendiendo hasta tener el máximo en torno a  $\beta = 1$ . La forma de ascender y descender varía en cada caballo, así como el caballo 1 desciende casi en línea recta, el caballo 2 lo hace en forma exponencial. Igualmente, el caballo 13 asciende en forma logarítmica y el 14 siguiendo casi la línea recta.

Teniendo en cuenta que el caballo ganador ha sido el número 3, vemos que la probabilidad mayor para este caballo se obtiene con  $\beta = 0$ .

Por lo tanto, en este caso la  $\beta$  óptima sería  $\beta = 0$ .

### 3.1.1.2. Estudio LOO

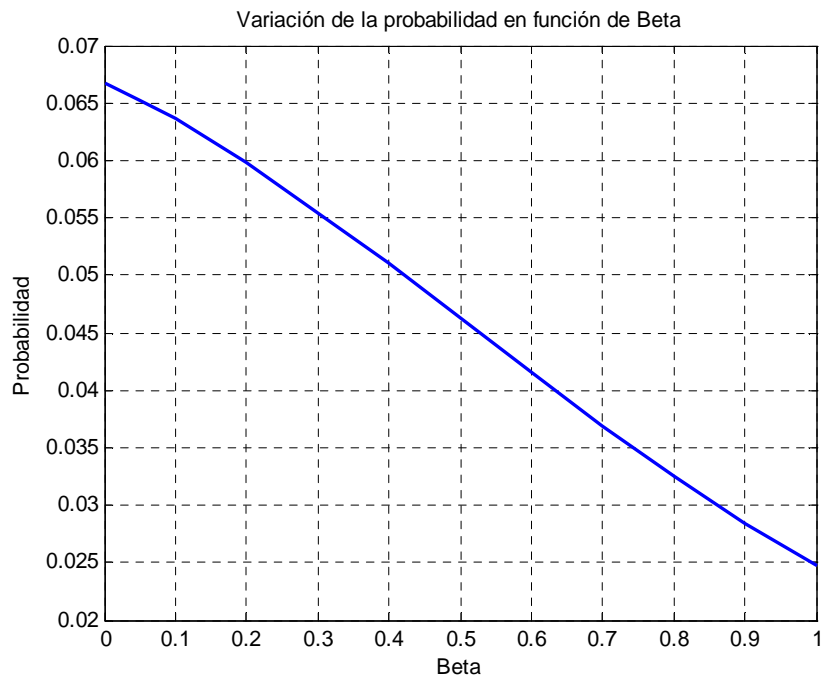
En este caso, vamos a usar la técnica de *Leave-One-Out*. Veremos cómo va variando la probabilidad en función del valor que elijamos de  $\beta$ .

	Beta=0	Beta=0,05	Beta=0,1	Beta=0,2	Beta=0,4	Beta=0,6	Beta=0,7	Beta=0,8	Beta=0,9	Beta=1
<b>Caballo 1</b>	0,0667	0,0652	0,0636	0,0599	0,0510	0,0415	0,0369	0,0325	0,0284	0,0247
<b>Caballo 2</b>	0,0667	0,0624	0,0582	0,0501	0,0357	0,0243	0,0198	0,0159	0,0127	0,0101
<b>Caballo 3</b>	0,0667	0,0622	0,0578	0,0494	0,0347	0,0233	0,0188	0,0150	0,0119	0,0094
<b>Caballo 4</b>	0,0667	0,0677	0,0685	0,0693	0,0685	0,0646	0,0618	0,0586	0,0551	0,0515
<b>Caballo 5</b>	0,0667	0,0633	0,0599	0,0530	0,0401	0,0289	0,0242	0,0201	0,0165	0,0135
<b>Caballo 6</b>	0,0667	0,0604	0,0545	0,0439	0,0275	0,0164	0,0125	0,0094	0,0071	0,0053
<b>Caballo 7</b>	0,0667	0,0613	0,0561	0,0466	0,0309	0,0196	0,0154	0,0120	0,0092	0,0071
<b>Caballo 8</b>	0,0667	0,0667	0,0665	0,0655	0,0610	0,0543	0,0505	0,0465	0,0426	0,0386
<b>Caballo 9</b>	0,0667	0,0696	0,0725	0,0777	0,0860	0,0909	0,0921	0,0925	0,0921	0,0911
<b>Caballo 10</b>	0,0667	0,0667	0,0665	0,0654	0,0608	0,0541	0,0502	0,0462	0,0422	0,0383
<b>Caballo 11</b>	0,0667	0,0728	0,0791	0,0926	0,1222	0,1540	0,1703	0,1867	0,2030	0,2192
<b>Caballo 12</b>	0,0667	0,0701	0,0735	0,0800	0,0911	0,0991	0,1018	0,1037	0,1048	0,1052
<b>Caballo 13</b>	0,0667	0,0725	0,0786	0,0914	0,1190	0,1480	0,1626	0,1771	0,1913	0,2052
<b>Caballo 14</b>	0,0667	0,0693	0,0719	0,0764	0,0831	0,0864	0,0867	0,0863	0,0853	0,0836
<b>Caballo 15</b>	0,0667	0,0699	0,0729	0,0787	0,0883	0,0945	0,0964	0,0974	0,0977	0,0972

**Tabla 3.2.** Ejemplo de las probabilidades de ganar para cada caballo por cada  $\beta$  con *Leave-One-Out*

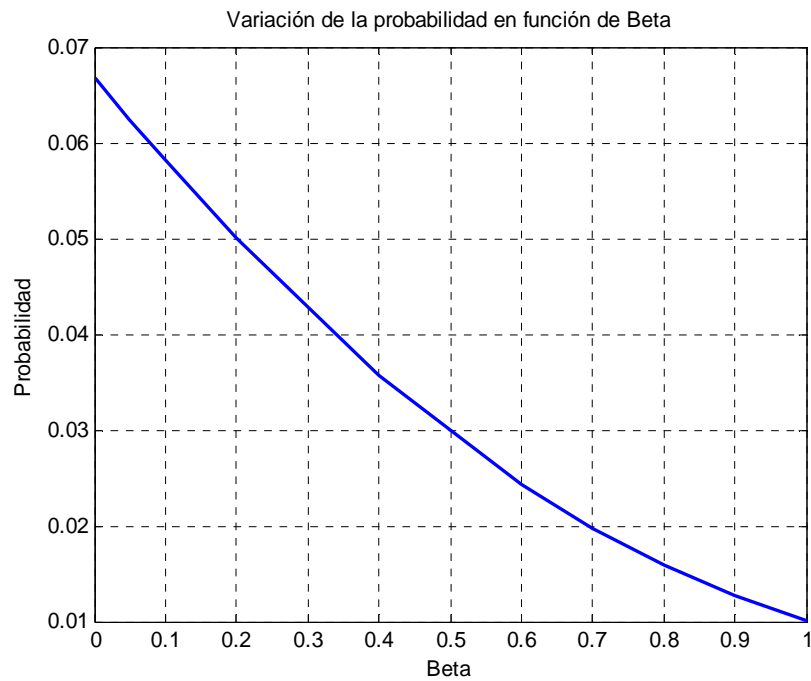
La representación de la probabilidad de ganar de cada caballo en función de  $\beta$  quedaría de la siguiente forma:

- Caballo 1



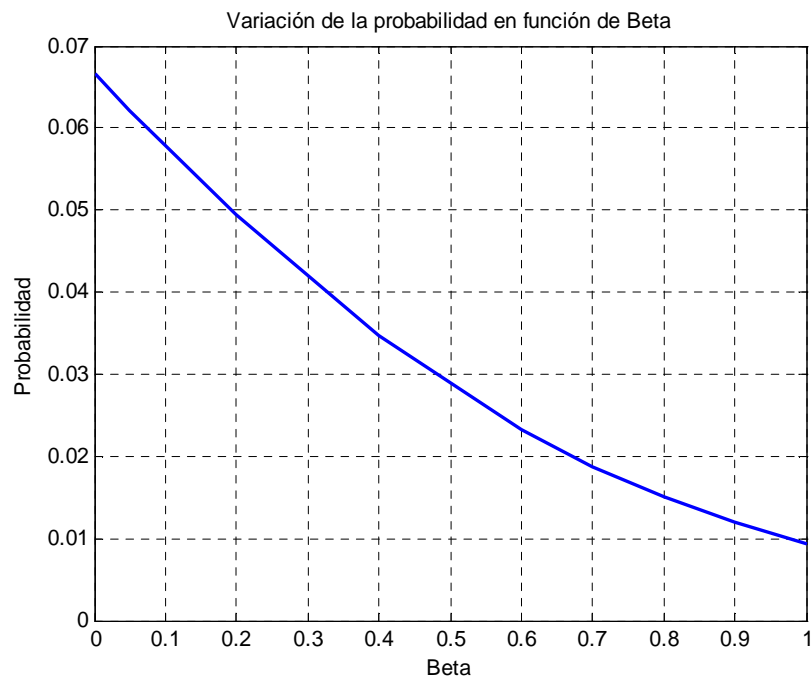
**Figura 3.16.** Variación de la probabilidad de ganar del caballo 1 en función de  $\beta$  con *Leave-One-Out*

- Caballo 2



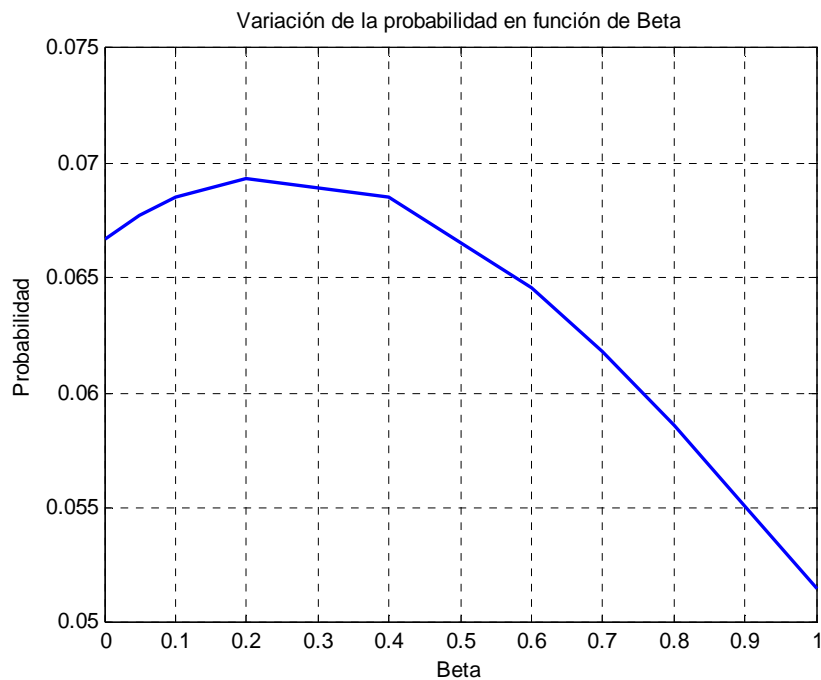
**Figura 3.17.** Variación de la probabilidad de ganar del caballo 2 en función de  $\beta$  con *Leave-One-Out*

- Caballo 3



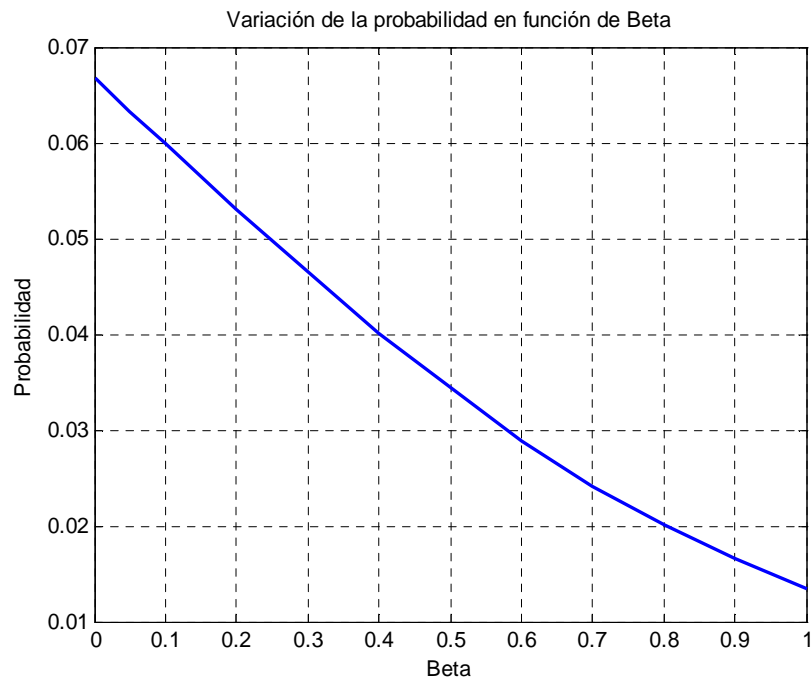
**Figura 3.18.** Variación de la probabilidad de ganar del caballo 3 en función de  $\beta$  con *Leave-One-Out*

- Caballo 4



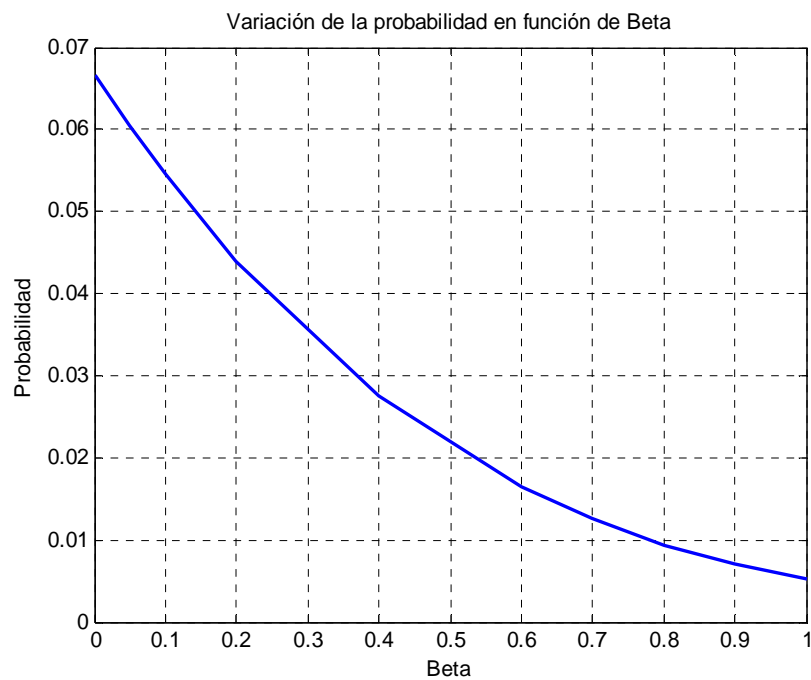
**Figura 3.19.** Variación de la probabilidad de ganar del caballo 4 en función de  $\beta$  con *Leave-One-Out*

- Caballo 5



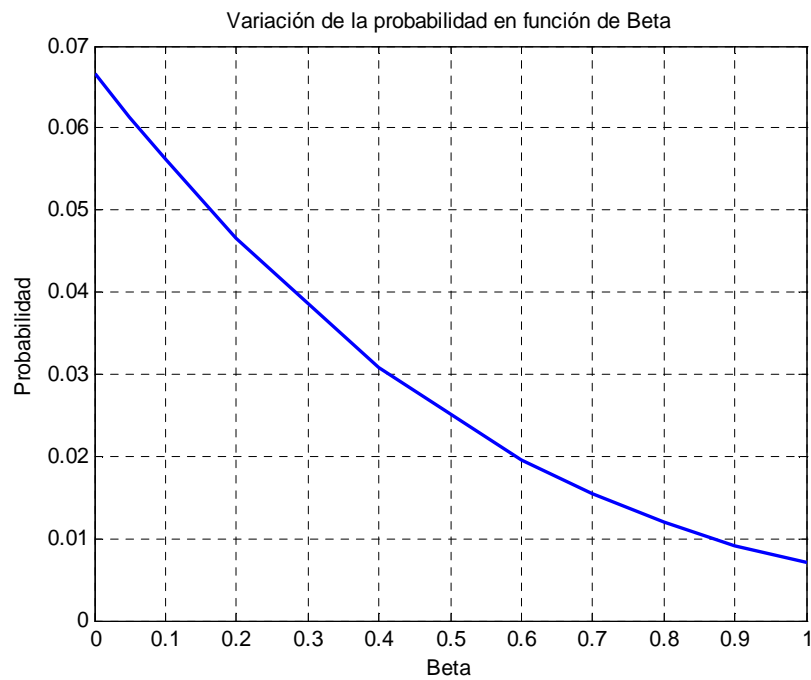
**Figura 3.20.** Variación de la probabilidad de ganar del caballo 5 en función de  $\beta$  con *Leave-One-Out*

- Caballo 6



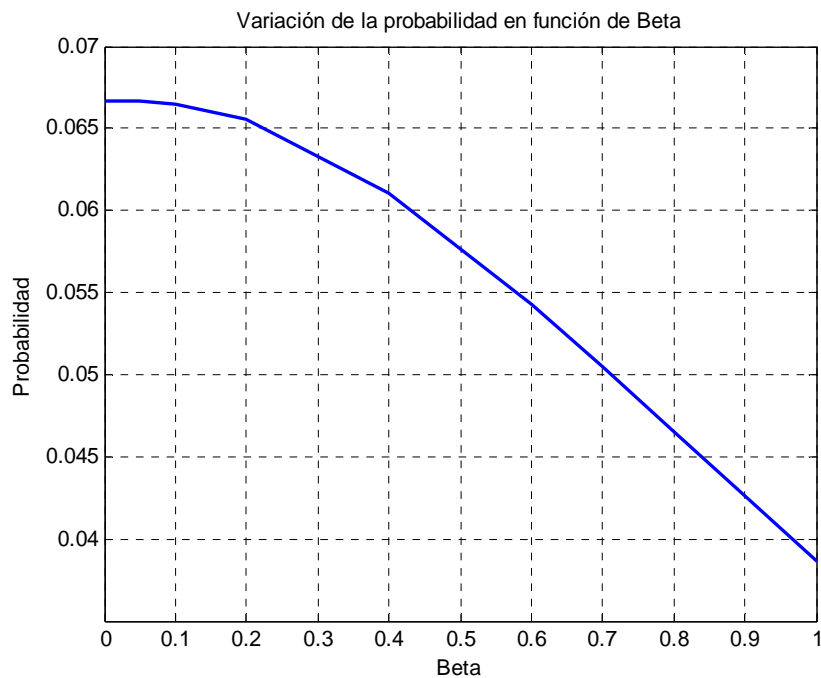
**Figura 3.21.** Variación de la probabilidad de ganar del caballo 6 en función de  $\beta$  con *Leave-One-Out*

- Caballo 7



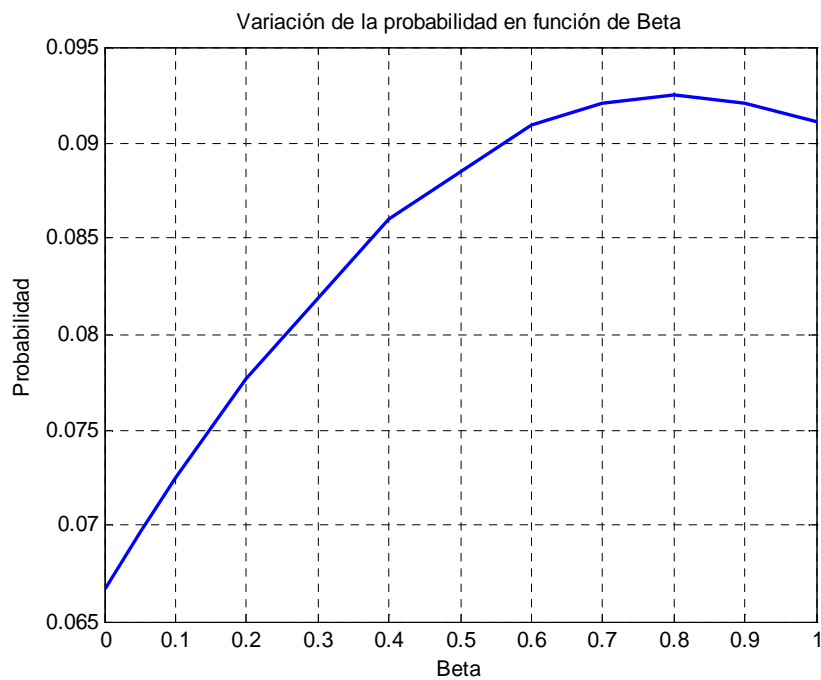
**Figura 3.22.** Variación de la probabilidad de ganar del caballo 7 en función de  $\beta$  con *Leave-One-Out*

- Caballo 8



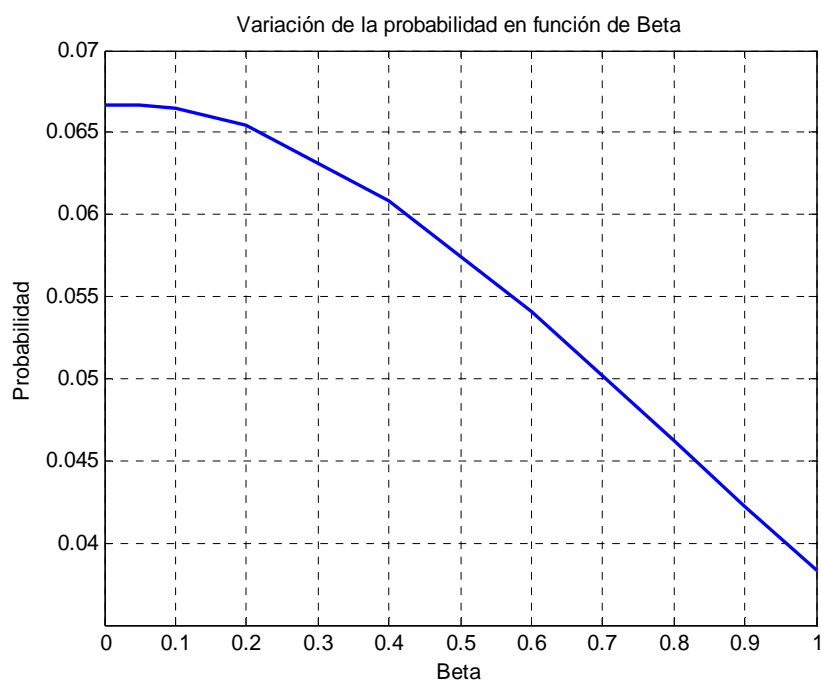
**Figura 3.23.** Variación de la probabilidad de ganar del caballo 8 en función de  $\beta$  con *Leave-One-Out*

- Caballo 9



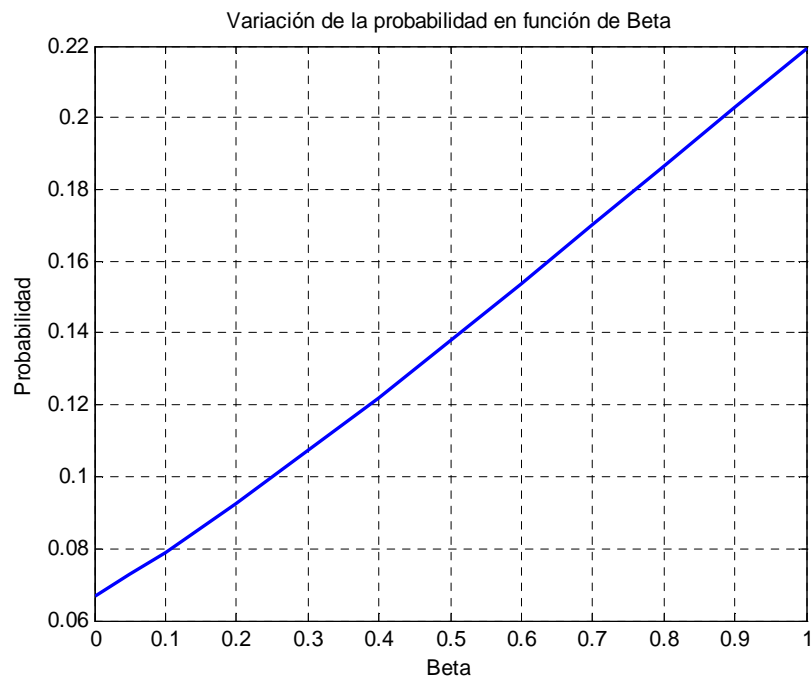
**Figura 3.24.** Variación de la probabilidad de ganar del caballo 9 en función de  $\beta$  con *Leave-One-Out*

- Caballo 10



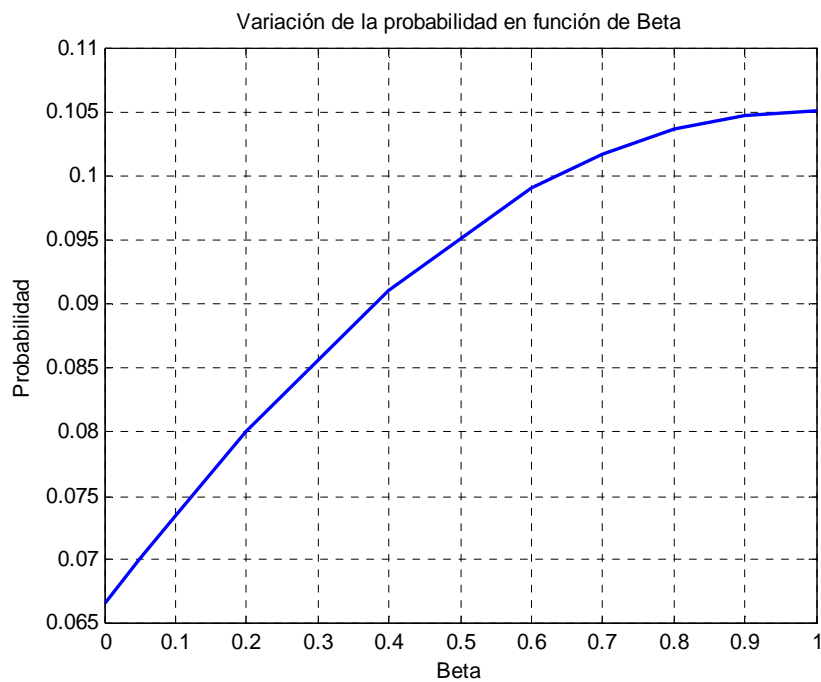
**Figura 3.25.** Variación de la probabilidad de ganar del caballo 10 en función de  $\beta$  con *Leave-One-Out*

- Caballo 11



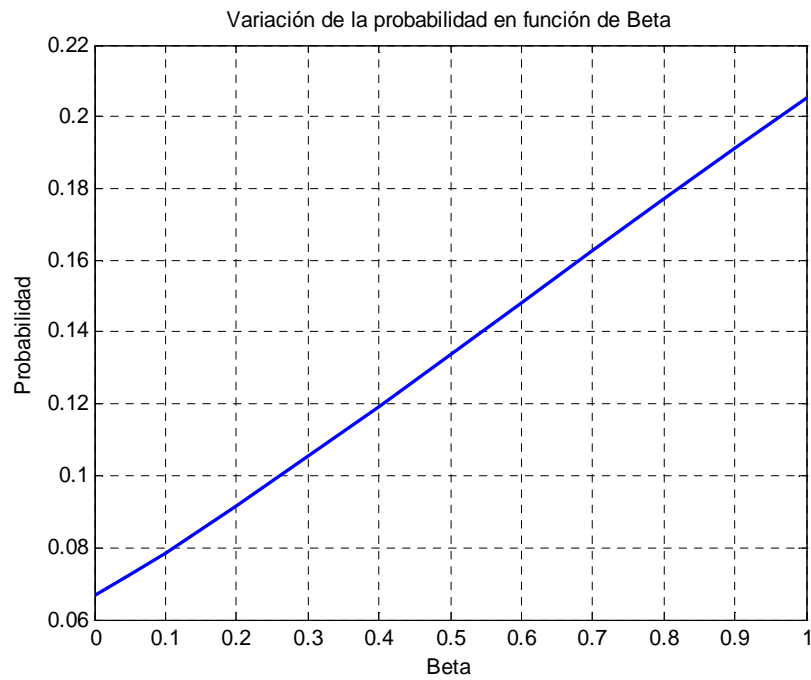
**Figura 3.26.** Variación de la probabilidad de ganar del caballo 11 en función de  $\beta$  con *Leave-One-Out*

- Caballo 12



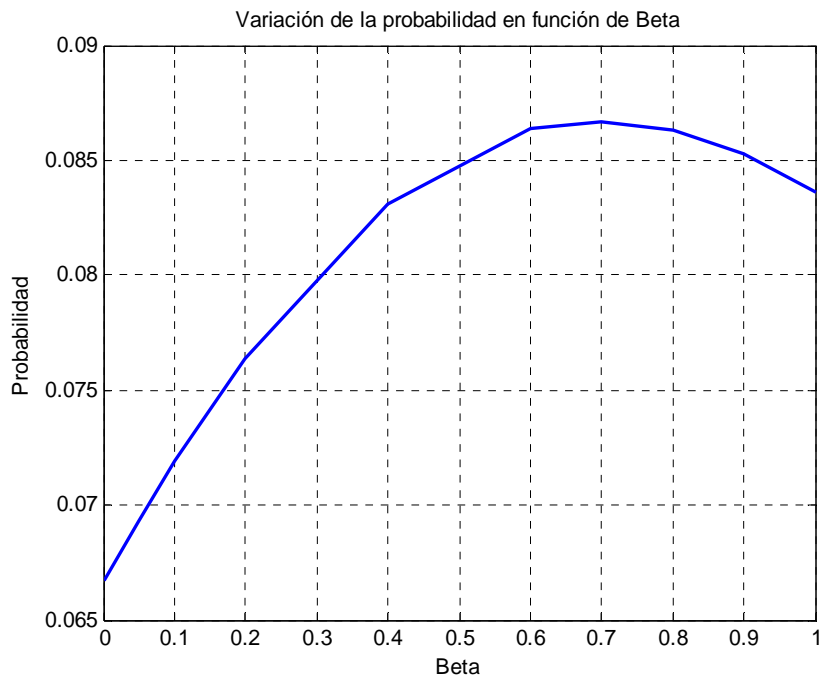
**Figura 3.27.** Variación de la probabilidad de ganar del caballo 12 en función de  $\beta$  con *Leave-One-Out*

- Caballo 13



**Figura 3.28.** Variación de la probabilidad de ganar del caballo 13 en función de  $\beta$  con *Leave-One-Out*

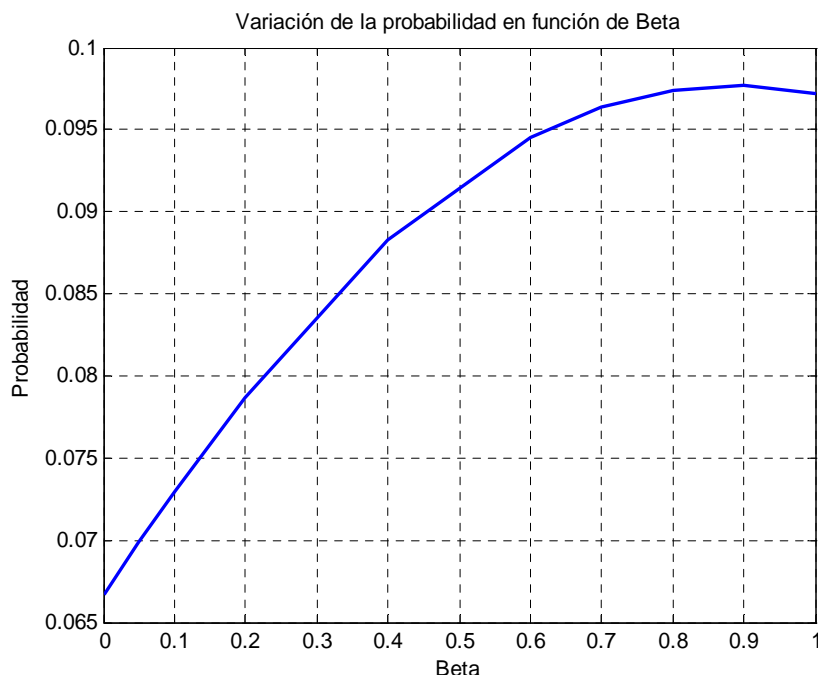
- Caballo 14



**Figura 3.29.** Variación de la probabilidad de ganar del caballo 14 en función de  $\beta$  con *Leave-One-Out*



- Caballo 15



**Figura 3.30.** Variación de la probabilidad de ganar del caballo 15 en función de  $\beta$  con *Leave-One-Out*

En este caso también podemos afirmar que cada caballo no varía su ganancia de igual forma, en función de  $\beta$ . En este caso, el caballo 1, caballo 2, caballo 3, caballo 5, caballo 6, caballo 7, caballo 8 y caballo 10 tienen su máximo con  $\beta = 0$  y desde ahí van bajando hasta llegar al mínimo en  $\beta = 1$ . Los caballos 4 y 14 empiezan en un valor alto con  $\beta = 0$ , pero, en lugar de ir descendiendo, suben teniendo su máximo en  $\beta = 0,2$  y  $\beta = 0,7$  respectivamente, para, a continuación, ir hasta su mínimo en  $\beta = 1$ . Y por último los caballos 9, 11, 12, 13 y 15 que en  $\beta = 0$  tienen su valor más pequeño y van ascendiendo hasta tener el máximo en torno a  $\beta = 1$ . La forma de ascender y descender varía en cada caballo, así como el caballo 1 desciende casi en línea recta, el caballo 2 lo hace en forma exponencial y en el caballo 8 en forma logarítmica. Igualmente, el caballo 9 asciende en forma logarítmica y el 11 siguiendo casi la línea recta.

Teniendo en cuenta que el caballo ganador ha sido el número 3, vemos que la probabilidad mayor para este caballo se obtiene con  $\beta = 0$ .

Por lo tanto, en este ejemplo la  $\beta$  óptima sería  $\beta = 0$ .

Además si lo comparamos con el resultado usando CV, vemos que es el mismo, por lo que, en este ejemplo, no podemos decidir entre cuál de los dos métodos nos lleva a un mejor resultado.

### **3.1.2. Variación de la ganancia en función de $\beta$**

En este apartado veremos cómo varía la ganancia, para cada carrera, en función de  $\beta$ , para así poder decidir qué valor de  $\beta$  es el mejor, a fin de obtener mayores beneficios.

Para ello vamos a realizar el estudio con cada uno de los métodos que hemos utilizado para realizar el entrenamiento de nuestro programa.

#### **3.1.2.1. Estudio CV**

La ganancia para cada carrera, en función de  $\beta$  será:

Carrera	Beta=0	Beta=0,05	Beta=0,1	Beta=0,2	Beta=0,4	Beta=0,6	Beta=0,7	Beta=0,8	Beta=0,9	Beta=1
1	-0,9414	-0,8974	-0,8533	-0,7651	-0,5888	-0,8701	-0,6973	-0,5246	-0,9636	-0,9101
2	-0,9368	-0,9126	-0,8884	-0,8401	-0,7433	-0,6465	-0,5981	-0,8961	-0,9118	-0,8017
3	-0,9418	-0,9173	-0,8929	-0,844	-0,7463	-0,6486	-0,9534	-0,8096	-0,8525	-0,7503
4	-0,9427	-0,8995	-0,8563	-0,7699	-0,5971	-0,8532	-0,6723	-0,8238	<b>2,7785</b>	<b>0,4709</b>
5	-0,942	-0,8971	-0,8523	-0,7626	-0,5833	-0,8338	-0,6525	-0,721	-0,8853	-0,3623
6	-0,9486	-0,9043	-0,8599	-0,7713	-0,5939	-0,8534	-0,6698	-0,7824	-0,8701	-0,191
7	-0,9442	-0,9035	-0,8628	-0,7814	-0,6187	-0,8581	-0,6777	<b>5,1565</b>	<b>3,4452</b>	<b>1,734</b>
8	-0,9204	-0,8844	-0,8484	-0,7765	-0,6325	-0,916	-0,7571	-0,8318	-0,9354	-0,4858
9	<b>46,3937</b>	<b>44,2908</b>	<b>42,1878</b>	<b>37,982</b>	<b>29,5701</b>	<b>21,1584</b>	<b>16,9524</b>	<b>12,7466</b>	<b>8,5407</b>	<b>4,3348</b>
10	-0,9369	-0,9064	-0,8759	-0,8149	-0,6929	-0,5708	-0,9123	-0,7601	-0,9851	-0,6412
11	-0,9527	-0,9251	-0,8974	-0,8422	-0,7317	-0,6212	-0,566	-0,9109	-0,7701	<b>0,5575</b>
12	-0,9322	-0,8938	-0,8554	-0,7785	-0,6248	-0,9388	-0,7799	-0,9163	-0,9951	-0,8808
13	-0,9373	-0,8944	-0,8514	-0,7655	-0,5938	-0,422	-0,9184	-0,7778	-0,9537	-0,9371
14	-0,925	-0,9057	-0,8863	-0,8476	-0,7702	-0,6928	-0,6541	-0,9026	-0,9415	-0,7361
15	-0,9178	-0,8832	-0,8486	-0,7793	-0,6408	-0,9698	-0,8159	-0,9112	-0,9368	-0,3907
16	-0,9353	-0,9037	-0,872	-0,8086	-0,6818	-0,9824	-0,8297	-0,7696	-0,9551	-0,2129
17	-0,9427	-0,9151	-0,8876	-0,8326	-0,7224	-0,6123	-0,913	-0,766	-0,8744	-0,7335
18	-0,9573	-0,9128	-0,8682	-0,7792	-0,6012	-0,9996	-0,8307	-0,6618	-0,9384	-0,0838
19	-0,9276	-0,8861	-0,8446	-0,7616	-0,5955	-0,9338	-0,7651	-0,8592	-0,976	-0,156
20	-0,9484	-0,9041	-0,8598	-0,7712	-0,594	-0,8581	-0,8138	-0,9356	-0,903	-0,8743
21	-0,937	-0,906	-0,8749	-0,8128	-0,6886	-0,5644	-0,8508	-0,9373	-0,9791	-0,8289
22	-0,9293	-0,887	-0,8446	-0,76	-0,5906	-0,8092	-0,9539	-0,6712	-0,9962	-0,5194
23	-0,9367	-0,9012	-0,8657	-0,7947	-0,6528	-0,5109	-0,8635	-0,9721	-0,9039	-0,9456
24	-0,9286	-0,884	-0,8393	-0,7501	-0,5717	-0,8204	-0,7626	-0,8613	-0,9601	-0,7933
25	-0,9335	-0,893	-0,8524	-0,7712	-0,6089	-0,8984	-0,7306	-0,8243	-0,9003	-0,9084
26	-0,9393	-0,9211	-0,903	-0,8666	-0,7939	-0,7213	-0,6849	-0,8866	-0,936	<b>0,3087</b>
27	-0,9462	-0,91	-0,8738	-0,8014	-0,6565	-0,9349	-0,7689	-0,8827	-0,8588	-0,7438
28	-0,9607	-0,9159	-0,871	-0,7813	-0,6019	-0,9487	-0,7808	<b>8,0238</b>	<b>4,2251</b>	<b>0,4264</b>
29	-0,9342	-0,8916	-0,8491	-0,7639	-0,5936	-0,8283	-0,6441	-0,7479	-0,9785	-0,2925
30	-0,9356	-0,9047	-0,8738	-0,8119	-0,6883	-0,5646	-0,8881	-0,7423	-0,8058	-0,9117
31	-0,9409	-0,9001	-0,8592	-0,7774	-0,6138	-0,9869	-0,8314	-0,9232	-0,8842	-0,9834
32	-0,9539	-0,9096	-0,8652	-0,7765	-0,5992	-0,9585	-0,7931	-0,6278	-0,931	-0,7913
33	-0,9146	-0,8815	-0,8484	-0,7821	-0,6496	-0,9575	-0,8059	-0,9222	-0,8846	-0,6743
34	-0,9409	-0,9084	-0,8759	-0,811	-0,6811	-0,9847	-0,8298	-0,9199	-0,9631	-0,642
35	-0,9446	-0,9088	-0,873	-0,8014	-0,6582	-0,9497	-0,779	-0,8616	-0,7729	-0,5815
36	-0,9557	-0,9127	-0,8697	-0,7837	-0,6117	-0,8926	-0,7139	-0,8895	-0,9814	<b>0,1755</b>
37	-0,9099	-0,8774	-0,8448	-0,7798	-0,6497	-0,9755	-0,8302	-0,6848	<b>13,332</b>	<b>9,7608</b>
38	-0,9515	-0,9081	-0,8647	-0,7779	-0,6044	-0,6019	-0,903	-0,7553	-0,8722	-0,9372
39	-0,9354	-0,9058	-0,8762	-0,817	-0,6987	-0,5803	-0,8489	-0,9548	-0,9258	-0,2587
40	-0,9562	-0,9189	-0,8816	-0,8071	-0,6579	-0,5087	-0,8923	-0,741	-0,9857	-0,8638

**Tabla 3.3.** Ejemplo de la ganancia de cada carrera en función de  $\hat{a}$  con validación cruzada

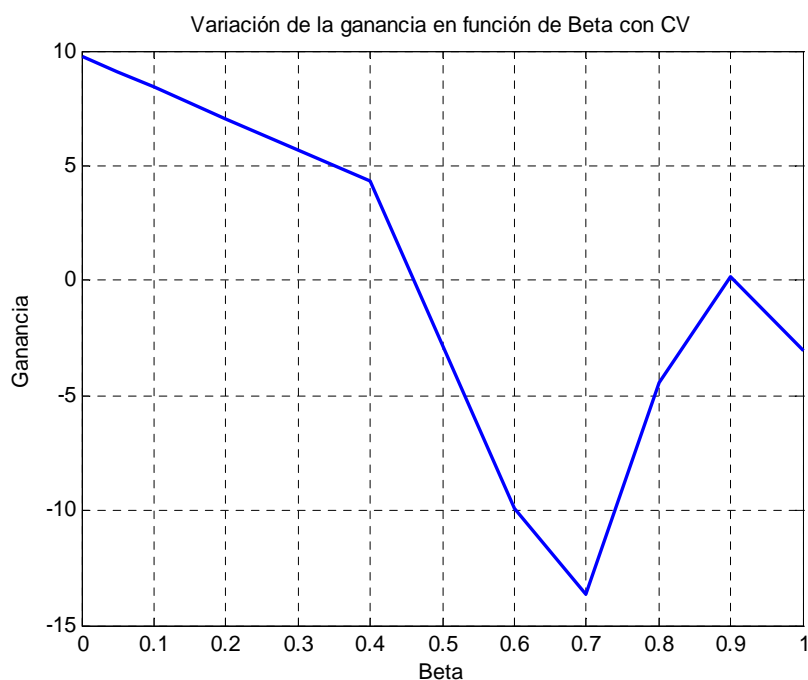
Como vemos, excepto para unos pocos valores (en negrita), el resto son todo pérdidas.

Podemos hacer una tabla resumen de lo anterior, con la ganancia total para todas las carreras por las que hemos apostado, y su variación con  $\beta$ :

Beta	0	0,05	0,1	0,2	0,4	0,6	0,7	0,8	0,9	1
Ganancia	9,7779	9,0985	8,42	7,0621	4,346	-9,9203	-13,6804	-4,4393	0,154	-3,0548

**Tabla 3.4.** Ganancia total para cada  $\hat{a}$  con validación cruzada

Realizamos una representación gráfica para verlo más claramente.



**Figura 3.31.** Variación de la ganancia total en función de  $\hat{a}$  con validación cruzada

Como podemos apreciar, el beneficio más alto lo alcanzaríamos con  $\beta = 0$ . Con este valor obtendríamos una ganancia de 9,7779€. Si tenemos en cuenta que apostamos 1€ por carrera, el beneficio total será de 49,7779€

### 3.1.2.2. Estudio LOO

Vamos a usar el mismo procedimiento para el caso de *Leave-One-Out*.

En primer lugar mostraremos la tabla con la ganancia que obtenemos para cada carrera.

Carrera	Beta=0	Beta=0,05	Beta=0,1	Beta=0,2	Beta=0,4	Beta=0,6	Beta=0,7	Beta=0,8	Beta=0,9	Beta=1
1	-0,9414	-0,8992	-0,8569	-0,7724	-0,6034	-0,9383	-0,7769	-0,8862	-0,9362	-0,9281
2	-0,9368	-0,9116	-0,8863	-0,8357	-0,7346	-0,6335	-0,583	-0,8985	-0,9181	-0,8138
3	-0,9418	-0,9099	-0,878	-0,8143	-0,6868	-0,5593	-0,8824	-0,7283	-0,7776	-0,6311
4	-0,9427	-0,9007	-0,8587	-0,7747	-0,6066	-0,8625	-0,6831	<b>5,7339</b>	<b>3,5073</b>	<b>1,2805</b>
5	-0,942	-0,9013	-0,8606	-0,7792	-0,6163	-0,9036	-0,7339	-0,8282	-0,9811	-0,5649
6	-0,9486	-0,9053	-0,8621	-0,7756	-0,6026	-0,8644	-0,6827	-0,8259	-0,9442	-0,0638
7	-0,9442	-0,9082	-0,8722	-0,8001	-0,656	-0,9158	-0,7451	<b>4,9947</b>	<b>3,2633</b>	<b>1,5319</b>
8	-0,9204	-0,8827	-0,8451	-0,7697	-0,6189	-0,9457	-0,7918	-0,8721	-0,9682	-0,0997
9	<b>46,3937</b>	<b>44,3988</b>	<b>42,4039</b>	<b>38,4142</b>	<b>30,4347</b>	<b>22,4552</b>	<b>18,4654</b>	<b>14,4757</b>	<b>10,4859</b>	<b>6,4962</b>
10	-0,9369	-0,9001	-0,8632	-0,7895	-0,642	-0,9631	-0,7939	-0,9054	-0,8022	-0,7609
11	-0,9527	-0,9316	-0,9106	-0,8686	-0,7845	-0,7003	-0,985	-0,8381	-0,915	0,0963
12	-0,9322	-0,8955	-0,8588	-0,7854	-0,6385	-0,9397	-0,7808	-0,915	-0,9394	-0,858
13	-0,9373	-0,8956	-0,8538	-0,7704	-0,6035	-0,4366	-0,8733	-0,9631	-0,9099	-0,8544
14	-0,925	-0,8986	-0,8723	-0,8196	-0,7141	-0,6087	-0,9384	-0,7971	-0,8911	-0,65
15	-0,9178	-0,8794	-0,841	-0,7642	-0,6105	-0,9039	-0,739	-0,8566	-0,8944	-0,4985
16	-0,9353	-0,9014	-0,8676	-0,7998	-0,6642	-0,9563	-0,7993	-0,9781	-0,937	0,0344
17	-0,9427	-0,9148	-0,8869	-0,8312	-0,7198	-0,6083	-0,9359	-0,7921	-0,9136	-0,7495
18	-0,9573	-0,9129	-0,8685	-0,7797	-0,6021	-0,984	-0,8125	-0,9654	-0,9119	-0,1893
19	-0,9276	-0,8897	-0,8518	-0,7759	-0,6242	-0,9896	-0,8302	-0,9697	-0,957	-0,4316
20	-0,9484	-0,9028	-0,8573	-0,7662	-0,854	-0,8687	-0,6848	-0,8774	-0,8358	-0,9541
21	-0,937	-0,905	-0,8731	-0,8091	-0,6812	-0,9869	-0,8335	-0,9174	-0,847	-0,7209
22	-0,9283	-0,8871	-0,8449	-0,7604	-0,5916	-0,8146	-0,9717	-0,6915	-0,9137	-0,4738
23	-0,9367	-0,9067	-0,8768	-0,817	-0,6973	-0,5776	-0,9227	-0,7801	-0,9331	-0,9605
24	-0,9286	-0,8832	-0,8378	-0,7471	-0,5655	-0,9519	-0,789	-0,878	-0,7694	-0,7512
25	-0,9335	-0,8926	-0,8517	-0,7698	-0,606	-0,8912	-0,7222	-0,809	-0,9114	-0,8772
26	-0,9393	-0,9259	-0,9125	-0,8857	-0,8321	-0,7785	-0,7517	-0,9644	-0,8441	<b>0,1313</b>
27	-0,9462	-0,9059	-0,8657	-0,7851	-0,6241	-0,8898	-0,7163	-0,8232	-0,8212	-0,631
28	-0,9607	-0,917	-0,8734	-0,786	-0,6113	-0,4367	-0,8579	<b>8,188</b>	<b>4,4098</b>	<b>0,6317</b>
29	-0,9342	-0,8919	-0,8495	-0,7649	-0,5956	-0,8396	-0,6573	-0,7494	-0,7613	-0,4039
30	-0,9356	-0,907	-0,8784	-0,8212	-0,7068	-0,5924	-0,884	-0,7376	-0,8083	-0,8764
31	-0,9409	-0,8998	-0,8587	-0,7764	-0,6119	-0,951	-0,7895	-0,9206	-0,8645	-0,9831
32	-0,9539	-0,9089	-0,8639	-0,7738	-0,5937	-0,9548	-0,7889	-0,9539	-0,9627	-0,7789
33	-0,9146	-0,8845	-0,8544	-0,7942	-0,6738	-0,9468	-0,8775	-0,9546	-0,9401	-0,313
34	-0,9409	-0,9051	-0,8692	-0,7976	-0,6544	-0,9596	-0,8005	-0,8876	-0,9495	-0,6117
35	-0,9446	-0,916	-0,8873	-0,8301	-0,7156	-0,6011	-0,9222	-0,771	-0,8319	-0,6596
36	-0,9557	-0,9137	-0,8716	-0,7875	-0,6194	-0,8731	-0,6911	-0,8245	-0,9596	0,1049
37	-0,9099	-0,8738	-0,8377	-0,7656	-0,6212	-0,9448	-0,7943	<b>14,5699</b>	<b>10,707</b>	<b>6,8439</b>
38	-0,9515	-0,9095	-0,8674	-0,7833	-0,6151	-0,9508	-0,7865	-0,9597	-0,9402	-0,911
39	-0,9354	-0,901	-0,8665	-0,7977	-0,6599	-0,9459	-0,7853	-0,9159	-0,8737	-0,4256
40	-0,9562	-0,924	-0,8918	-0,8273	-0,6983	-0,5694	-0,9576	-0,8156	-0,9969	-0,908

**Tabla 3.5.** Ejemplo de la ganancia de cada carrera en función de  $\hat{a}$  con *Leave-One-Out*

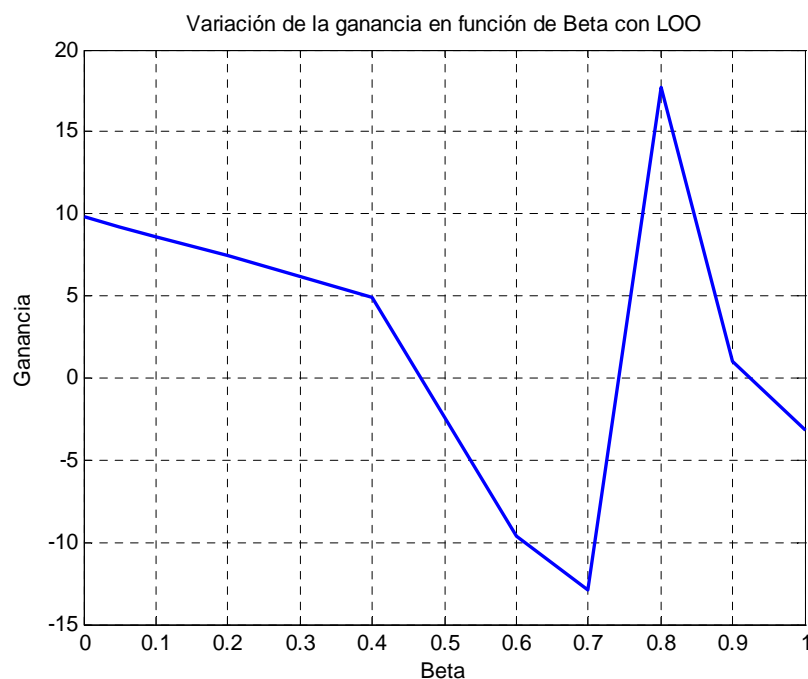
En este caso, igual que en el anterior, para casi todas las carreras tenemos pérdida de dinero. En las únicas en las que ganamos son las mostradas en negrita.

Como resumen diremos que:

Beta	0	0,05	0,1	0,2	0,4	0,6	0,7	0,8	0,9	1
Ganancia	9,7789	9,1989	8,6199	7,4622	4,8773	-9,5836	-12,8863	17,711	1,012	-3,1824

**Tabla 3.6.** Ganancia total para cada  $\hat{a}$  con *Leave-One-Out*

Gráficamente podemos verlo como:



**Figura 3.32.** Variación de la ganancia total en función de  $\hat{a}$  con *Leave-One-Out*

En este caso, el valor óptimo de  $\beta$  para maximizar la ganancia, se daría con  $\beta = 0,8$ , donde tenemos un beneficio de 17,711€. Por lo tanto si tenemos en cuenta que hemos apostado 1€ por carrera, terminaremos nuestras apuestas con un total de 57,711€.

Por lo tanto, el mejor resultado lo vamos a tener con  $\beta = 0,8$  y el método de LOO.

### 3.1.3. Elección del valor óptimo de $\beta$

Según hemos visto en los apartados anteriores, en función de la ganancia podríamos decir que el mejor valor de  $\beta$  es  $\beta = 0,8$ , y, además, con el método de LOO obtenemos resultados superiores que con CV.

Sin embargo para sacar una conclusión final, vamos a centrarnos más en las apuestas que realizamos a través del Criterio de Kelly.

Para ello vamos a escoger los parámetros  $c = 25$  y  $\sigma = 15$ , que son los que hemos visto que nos reportaban unos mejores resultados.

En primer lugar, vamos a mostrar las tablas donde recogemos:

- Cantidad total apostada en cada carrera, es decir, la suma de todo lo que apostamos por cada caballo en una carrera.
- La ganancia que obtenemos en cada carrera.

A continuación, expondremos los vectores obtenidos con el Criterio de Kelly, de nombre “ $a$ ”, para el parámetro  $\beta$ . Así iremos observando cómo varía el número de caballos por los que vamos apostando en cada carrera, en función de dicho parámetro.

La forma en la que vamos a presentarlo es, para cada tipo de validación, primero mostraremos dos tablas, de las citadas anteriormente, cada una con un valor de  $\beta$  distinto. Y, seguidamente, pondremos los vectores de las apuestas de Kelly pertenecientes a cada tabla.

#### **3.1.3.1. Estudio CV**

$\beta = 0$ 

Tipo	Apuestas totales	Ganancia
1	0,9414	0
2	0,9368	0
3	0,9418	0
4	0,9427	0
5	0,942	0
6	0,9486	0
7	0,9442	0
8	0,9204	0
9	0,9279	47,3216
10	0,9369	0
11	0,9527	0
12	0,9322	0
13	0,9373	0
14	0,925	0
15	0,9178	0
16	0,9353	0
17	0,9427	0
18	0,9573	0
19	0,9276	0
20	0,9484	0
21	0,937	0
22	0,9293	0
23	0,9367	0
24	0,9286	0
25	0,9335	0
26	0,9393	0
27	0,9462	0
28	0,9607	0
29	0,9342	0
30	0,9356	0
31	0,9409	0
32	0,9539	0
33	0,9146	0
34	0,9409	0
35	0,9446	0
36	0,9557	0
37	0,9099	0
38	0,9515	0
39	0,9354	0
40	0,9562	0
<b>Total</b>	<b>37,5437</b>	<b>47,3216</b>

 $\beta = 0,05$ 

Tipo	Apuestas totales	Ganancia
1	0,8974	0
2	0,9126	0
3	0,9173	0
4	0,8995	0
5	0,8971	0
6	0,9043	0
7	0,9035	0
8	0,8844	0
9	0,8858	45,1766
10	0,9064	0
11	0,9251	0
12	0,8938	0
13	0,8944	0
14	0,9057	0
15	0,8832	0
16	0,9037	0
17	0,9151	0
18	0,9128	0
19	0,8861	0
20	0,9041	0
21	0,906	0
22	0,887	0
23	0,9012	0
24	0,884	0
25	0,893	0
26	0,9211	0
27	0,91	0
28	0,9159	0
29	0,8916	0
30	0,9047	0
31	0,9001	0
32	0,9096	0
33	0,8815	0
34	0,9084	0
35	0,9088	0
36	0,9127	0
37	0,8774	0
38	0,9081	0
39	0,9058	0
40	0,9189	0
<b>Total</b>	<b>36,0781</b>	<b>45,1766</b>

**Tabla 3.7.** Validación cruzada: apuestas totales y ganancias para cada tipo, con  $\hat{a} = 0$  y  $\hat{a} = 0,05$



- $\beta = 0$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0	0,9414	0	0	0	
a2	0	0	0	0	0	0	0	0	0	0,9368						
a3	0	0	0	0	0	0	0	0	0,9418	0						
a4	0	0	0	0	0	0	0	0	0	0,9427	0	0	0			
a5	0	0	0	0	0	0	0	0	0	0	0	0	0	0,942		
a6	0	0	0	0	0	0	0	0	0	0	0	0,9486	0	0		
a7	0	0	0	0	0	0	0	0	0	0	0	0	0,9442			
a8	0	0	0	0	0	0	0	0	0	0	0	0	0,9204	0		
a9	0	0	0	0	0	0	0	0	0	0	0	0	0	0,9279		
a10	0	0	0,9369	0	0	0	0	0	0	0						
a11	0	0	0	0	0	0	0	0	0	0,9527						
a12	0	0	0	0	0	0	0	0	0	0	0	0	0,9322	0		
a13	0	0	0	0	0	0	0	0	0	0	0	0	0,9373	0		
a14	0	0	0	0	0	0	0	0	0	0	0,925					
a15	0	0	0	0	0	0	0	0,9178	0	0	0	0	0	0		
a16	0	0	0	0	0	0	0	0	0	0	0	0	0,9353			
a17	0	0	0	0	0	0	0	0	0	0	0,9427	0	0			
a18	0	0	0	0	0	0	0	0	0	0,9573						
a19	0	0	0	0	0	0	0	0	0	0,9276						
a20	0,9484	0	0	0	0	0	0	0	0	0						
a21	0	0	0	0	0	0	0	0	0	0	0	0	0,937	0		
a22	0	0	0	0	0	0	0	0	0,9293	0	0	0	0	0	0	0
a23	0	0	0	0	0	0	0	0	0,9367	0	0					
a24	0	0	0	0	0	0	0,9286	0	0	0	0	0	0	0		
a25	0	0,9335	0	0	0	0	0	0	0	0	0	0	0	0		
a26	0	0	0	0	0	0	0	0	0	0,9393						
a27	0	0	0	0	0	0	0	0,9462	0	0	0	0	0	0		
a28	0	0	0	0	0	0	0	0	0	0,9607	0	0	0	0		
a29	0	0	0	0	0	0	0	0	0	0	0,9342	0				
a30	0	0	0	0	0	0	0	0,9356	0	0						
a31	0	0	0	0	0	0	0	0	0	0,9409	0	0	0	0		
a32	0	0	0	0	0	0	0	0	0,9539	0	0	0				
a33	0	0	0	0	0	0	0,9146	0	0	0	0	0	0	0	0	0
a34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,9409
a35	0	0	0	0	0	0	0	0	0,9446	0						
a36	0	0	0	0	0	0	0	0	0	0,9557						
a37	0	0	0	0	0	0	0	0	0,9099	0	0	0	0	0		
a38	0	0	0	0	0	0	0	0	0	0,9515	0	0	0	0		
a39	0	0	0	0	0	0	0	0	0	0	0,9354	0	0	0		
a40	0	0	0	0	0	0	0	0	0,9562							

**Tabla 3.8.** Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0$

- $\beta = 0,05$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0	0,8974	0	0	0	0
a2	0	0	0	0	0	0	0	0	0	0,9126						
a3	0	0	0	0	0	0	0	0	0,9173	0						
a4	0	0	0	0	0	0	0	0	0	0,8995	0	0	0			
a5	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8971		
a6	0	0	0	0	0	0	0	0	0	0	0	0,9043	0	0		
a7	0	0	0	0	0	0	0	0	0	0	0	0	0,9035			
a8	0	0	0	0	0	0	0	0	0	0	0	0	0,8844	0		
a9	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8858		
a10	0	0	0	0	0	0	0	0	0	0,9064						
a11	0	0	0	0	0	0	0	0	0	0,9251						
a12	0	0	0	0	0	0	0	0	0	0	0	0	0,8938	0		
a13	0	0	0	0	0	0	0	0	0	0	0	0	0,8944	0		
a14	0	0	0	0	0	0	0	0	0	0	0,9057					
a15	0	0	0	0	0	0	0	0,8832	0	0	0	0	0	0	0	
a16	0	0	0	0	0	0	0	0	0	0	0	0	0	0,9037		
a17	0	0	0	0	0	0	0	0	0	0	0,9151	0	0			
a18	0	0	0	0	0	0	0	0	0	0,9128						
a19	0	0	0	0	0	0	0	0	0	0,8861						
a20	0,9041	0	0	0	0	0	0	0	0	0						
a21	0	0	0	0	0	0	0	0	0	0	0	0	0,906	0		
a22	0	0	0	0	0	0	0	0	0,887	0	0	0	0	0	0	0
a23	0	0	0	0	0	0	0	0	0,9012	0	0					
a24	0	0	0	0	0	0	0,884	0	0	0	0	0	0	0	0	
a25	0	0,893	0	0	0	0	0	0	0	0	0	0	0	0	0	
a26	0	0	0	0	0	0	0	0	0	0,9211						
a27	0	0	0	0	0	0	0	0,91	0	0	0	0	0	0	0	
a28	0	0	0	0	0	0	0	0	0	0,9159	0	0	0	0	0	
a29	0	0	0	0	0	0	0	0	0	0	0,8916	0				
a30	0	0	0	0	0	0	0	0,9047	0	0						
a31	0	0	0	0	0	0	0	0	0	0,9001	0	0	0	0		
a32	0	0	0	0	0	0	0	0	0,9096	0	0	0				
a33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8815	
a34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,9084
a35	0	0	0	0	0	0	0	0	0,9088	0						
a36	0	0	0	0	0	0	0	0	0	0,9127						
a37	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8774		
a38	0	0	0	0	0	0	0	0	0	0,9081	0	0	0	0		
a39	0	0	0	0	0	0	0	0	0	0	0,9058	0	0	0		
a40	0	0	0	0	0	0	0	0	0,9189							

**Tabla 3.9.** Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,05$

$\beta = 0,1$ 

Tipo	Apuestas totales	Ganancia
1	0,8533	0
2	0,8884	0
3	0,8929	0
4	0,8563	0
5	0,8523	0
6	0,8599	0
7	0,8628	0
8	0,8484	0
9	0,8438	43,0316
10	0,8759	0
11	0,8974	0
12	0,8554	0
13	0,8514	0
14	0,8863	0
15	0,8486	0
16	0,872	0
17	0,8876	0
18	0,8682	0
19	0,8446	0
20	0,8598	0
21	0,8749	0
22	0,8446	0
23	0,8657	0
24	0,8393	0
25	0,8524	0
26	0,903	0
27	0,8738	0
28	0,871	0
29	0,8491	0
30	0,8738	0
31	0,8592	0
32	0,8652	0
33	0,8484	0
34	0,8759	0
35	0,873	0
36	0,8697	0
37	0,8448	0
38	0,8647	0
39	0,8762	0
40	0,8816	0
<b>Total</b>	<b>34,6116</b>	<b>43,0316</b>

 $\beta = 0,2$ 

Tipo	Apuestas totales	Ganancia
1	0,7651	0
2	0,8401	0
3	0,844	0
4	0,7699	0
5	0,7626	0
6	0,7713	0
7	0,7814	0
8	0,7765	0
9	0,7596	38,7416
10	0,8149	0
11	0,8422	0
12	0,7785	0
13	0,7655	0
14	0,8476	0
15	0,7793	0
16	0,8086	0
17	0,8326	0
18	0,7792	0
19	0,7616	0
20	0,7712	0
21	0,8128	0
22	0,76	0
23	0,7947	0
24	0,7501	0
25	0,7712	0
26	0,8666	0
27	0,8014	0
28	0,7813	0
29	0,7639	0
30	0,8119	0
31	0,7774	0
32	0,7765	0
33	0,7821	0
34	0,811	0
35	0,8014	0
36	0,7837	0
37	0,7798	0
38	0,7779	0
39	0,817	0
40	0,8071	0
<b>Total</b>	<b>31,6795</b>	<b>38,7416</b>

**Tabla 3.10.** Validación cruzada: apuestas totales y ganancias para cada tipo, con  $\beta = 0,1$  y  $\beta = 0,2$

- $\beta = 0,1$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0	0,8533	0	0	0	0
a2	0	0	0	0	0	0	0	0	0	0,8884						
a3	0	0	0	0	0	0	0	0	0,8929	0						
a4	0	0	0	0	0	0	0	0	0	0,8563	0	0	0			
a5	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8523		
a6	0	0	0	0	0	0	0	0	0	0	0	0,8599	0	0		
a7	0	0	0	0	0	0	0	0	0	0	0	0	0,8628			
a8	0	0	0	0	0	0	0	0	0	0	0	0	0,8484	0		
a9	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8438		
a10	0	0	0	0	0	0	0	0	0	0,8759						
a11	0	0	0	0	0	0	0	0	0	0,8974						
a12	0	0	0	0	0	0	0	0	0	0	0	0	0,8554	0		
a13	0	0	0	0	0	0	0	0	0	0	0	0	0,8514	0		
a14	0	0	0	0	0	0	0	0	0	0	0,8863					
a15	0	0	0	0	0	0	0	0,8486	0	0	0	0	0	0	0	
a16	0	0	0	0	0	0	0	0	0	0	0	0	0,872			
a17	0	0	0	0	0	0	0	0	0	0	0,8876	0	0			
a18	0	0	0	0	0	0	0	0	0	0,8682						
a19	0	0	0	0	0	0	0	0	0	0,8446						
a20	0,8598	0	0	0	0	0	0	0	0	0						
a21	0	0	0	0	0	0	0	0	0	0	0	0	0,8749	0		
a22	0	0	0	0	0	0	0	0	0,8446	0	0	0	0	0	0	0
a23	0	0	0	0	0	0	0	0	0,8657	0	0					
a24	0	0	0	0	0	0	0,8393	0	0	0	0	0	0	0	0	
a25	0	0,8524	0	0	0	0	0	0	0	0	0	0	0	0	0	
a26	0	0	0	0	0	0	0	0	0	0,903						
a27	0	0	0	0	0	0	0	0,8738	0	0	0	0	0	0	0	
a28	0	0	0	0	0	0	0	0	0	0,871	0	0	0	0	0	
a29	0	0	0	0	0	0	0	0	0	0	0,8491	0				
a30	0	0	0	0	0	0	0	0,8738	0	0						
a31	0	0	0	0	0	0	0	0	0	0,8592	0	0	0	0		
a32	0	0	0	0	0	0	0	0	0,8652	0	0	0				
a33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8484	
a34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8759
a35	0	0	0	0	0	0	0	0	0,873	0						
a36	0	0	0	0	0	0	0	0	0	0,8697						
a37	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8448		
a38	0	0	0	0	0	0	0	0	0	0,8647	0	0	0	0		
a39	0	0	0	0	0	0	0	0	0	0	0,8762	0	0	0		
a40	0	0	0	0	0	0	0	0	0,8816							

**Tabla 3.11.** Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,1$

- $\beta = 0,2$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a <sub>1</sub>	0	0	0	0	0	0	0	0	0	0	0	0,7651	0	0	0	0
a <sub>2</sub>	0	0	0	0	0	0	0	0	0	0,8401						
a <sub>3</sub>	0	0	0	0	0	0	0	0	0,844	0						
a <sub>4</sub>	0	0	0	0	0	0	0	0	0	0,7699	0	0	0			
a <sub>5</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7626		
a <sub>6</sub>	0	0	0	0	0	0	0	0	0	0	0	0,7713	0	0		
a <sub>7</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,7814			
a <sub>8</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,7765	0		
a <sub>9</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7596		
a <sub>10</sub>	0	0	0	0	0	0	0	0	0	0,8149						
a <sub>11</sub>	0	0	0	0	0	0	0	0	0	0,8422						
a <sub>12</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,7785	0		
a <sub>13</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,7655	0		
a <sub>14</sub>	0	0	0	0	0	0	0	0	0	0	0,8476					
a <sub>15</sub>	0	0	0	0	0	0	0	0,7793	0	0	0	0	0	0	0	
a <sub>16</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8086		
a <sub>17</sub>	0	0	0	0	0	0	0	0	0	0	0,8326	0	0			
a <sub>18</sub>	0	0	0	0	0	0	0	0	0	0,7792						
a <sub>19</sub>	0	0	0	0	0	0	0	0	0	0,7616						
a <sub>20</sub>	0,7712	0	0	0	0	0	0	0	0	0						
a <sub>21</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,8128	0		
a <sub>22</sub>	0	0	0	0	0	0	0	0	0,76	0	0	0	0	0	0	0
a <sub>23</sub>	0	0	0	0	0	0	0	0	0,7947	0	0					
a <sub>24</sub>	0	0	0	0	0	0	0,7501	0	0	0	0	0	0	0	0	
a <sub>25</sub>	0	0,7712	0	0	0	0	0	0	0	0	0	0	0	0	0	
a <sub>26</sub>	0	0	0	0	0	0	0	0	0	0,8666						
a <sub>27</sub>	0	0	0	0	0	0	0	0,8014	0	0	0	0	0	0	0	
a <sub>28</sub>	0	0	0	0	0	0	0	0	0	0,7813	0	0	0	0	0	
a <sub>29</sub>	0	0	0	0	0	0	0	0	0	0,7639	0					
a <sub>30</sub>	0	0	0	0	0	0	0	0,8119	0	0						
a <sub>31</sub>	0	0	0	0	0	0	0	0	0	0,7774	0	0	0	0		
a <sub>32</sub>	0	0	0	0	0	0	0	0	0,7765	0	0	0				
a <sub>33</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7821	
a <sub>34</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,811
a <sub>35</sub>	0	0	0	0	0	0	0	0	0,8014	0						
a <sub>36</sub>	0	0	0	0	0	0	0	0	0	0,7837						
a <sub>37</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7798		
a <sub>38</sub>	0	0	0	0	0	0	0	0	0	0,7779	0	0	0	0		
a <sub>39</sub>	0	0	0	0	0	0	0	0	0	0	0,817	0	0	0		
a <sub>40</sub>	0	0	0	0	0	0	0	0	0,8071							

**Tabla 3.12.** Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,2$

$\beta = 0,4$ 

Tipo	Apuestas totales	Ganancia
1	0,5888	0
2	0,7433	0
3	0,7463	0
4	0,5971	0
5	0,5833	0
6	0,5939	0
7	0,6187	0
8	0,6325	0
9	0,5914	30,1615
10	0,6929	0
11	0,7317	0
12	0,6248	0
13	0,5938	0
14	0,7702	0
15	0,6408	0
16	0,6818	0
17	0,7224	0
18	0,6012	0
19	0,5955	0
20	0,594	0
21	0,6886	0
22	0,5906	0
23	0,6528	0
24	0,5717	0
25	0,6089	0
26	0,7939	0
27	0,6565	0
28	0,6019	0
29	0,5936	0
30	0,6883	0
31	0,6138	0
32	0,5992	0
33	0,6496	0
34	0,6811	0
35	0,6582	0
36	0,6117	0
37	0,6497	0
38	0,6044	0
39	0,6987	0
40	0,6579	0
Total	25,8155	30,1615

 $\beta = 0,6$ 

Tipo	Apuestas totales	Ganancia
1	0,8701	0
2	0,6465	0
3	0,6486	0
4	0,8532	0
5	0,8338	0
6	0,8534	0
7	0,8581	0
8	0,916	0
9	0,8658	22,0242
10	0,5708	0
11	0,6212	0
12	0,9388	0
13	0,422	0
14	0,6928	0
15	0,9698	0
16	0,9824	0
17	0,6123	0
18	0,9996	0
19	0,9338	0
20	0,8581	0
21	0,5644	0
22	0,8092	0
23	0,5109	0
24	0,8204	0
25	0,8984	0
26	0,7213	0
27	0,9349	0
28	0,9487	0
29	0,8283	0
30	0,5646	0
31	0,9869	0
32	0,9585	0
33	0,9575	0
34	0,9847	0
35	0,9497	0
36	0,8926	0
37	0,9755	0
38	0,6019	0
39	0,5803	0
40	0,5087	0
Total	31,9445	22,0242

Tabla 3.13. Validación cruzada: apuestas totales y ganancias para cada tipo, con  $\beta = 0,4$  y  $\beta = 0,6$

- $\beta = 0,4$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0	0,5888	0	0	0	0
a2	0	0	0	0	0	0	0	0	0	0,7433	0	0	0	0	0	0
a3	0	0	0	0	0	0	0	0	0,7463	0	0	0	0	0	0	0
a4	0	0	0	0	0	0	0	0	0	0,5971	0	0	0	0	0	0
a5	0	0	0	0	0	0	0	0	0	0	0	0	0	0,5833	0	0
a6	0	0	0	0	0	0	0	0	0	0	0	0,5939	0	0	0	0
a7	0	0	0	0	0	0	0	0	0	0	0	0	0,6187	0	0	0
a8	0	0	0	0	0	0	0	0	0	0	0	0	0,6325	0	0	0
a9	0	0	0	0	0	0	0	0	0	0	0	0	0	0,5914	0	0
a10	0	0	0	0	0	0	0	0	0	0,6929	0	0	0	0	0	0
a11	0	0	0	0	0	0	0	0	0	0,7317	0	0	0	0	0	0
a12	0	0	0	0	0	0	0	0	0	0	0	0	0,6248	0	0	0
a13	0	0	0	0	0	0	0	0	0	0	0	0	0,5938	0	0	0
a14	0	0	0	0	0	0	0	0	0	0	0,7702	0	0	0	0	0
a15	0	0	0	0	0	0	0	0,6408	0	0	0	0	0	0	0	0
a16	0	0	0	0	0	0	0	0	0	0	0	0	0,6818	0	0	0
a17	0	0	0	0	0	0	0	0	0	0	0,7224	0	0	0	0	0
a18	0	0	0	0	0	0	0	0	0	0,6012	0	0	0	0	0	0
a19	0	0	0	0	0	0	0	0	0	0,5955	0	0	0	0	0	0
a20	0,594	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a21	0	0	0	0	0	0	0	0	0	0	0	0	0,6886	0	0	0
a22	0	0	0	0	0	0	0	0	0,5906	0	0	0	0	0	0	0
a23	0	0	0	0	0	0	0	0	0,6528	0	0	0	0	0	0	0
a24	0	0	0	0	0	0	0,5717	0	0	0	0	0	0	0	0	0
a25	0	0,6089	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a26	0	0	0	0	0	0	0	0	0	0,7939	0	0	0	0	0	0
a27	0	0	0	0	0	0	0	0,6565	0	0	0	0	0	0	0	0
a28	0	0	0	0	0	0	0	0	0	0,6019	0	0	0	0	0	0
a29	0	0	0	0	0	0	0	0	0	0	0,5936	0	0	0	0	0
a30	0	0	0	0	0	0	0	0,6883	0	0	0	0	0	0	0	0
a31	0	0	0	0	0	0	0	0	0	0,6138	0	0	0	0	0	0
a32	0	0	0	0	0	0	0	0	0,5992	0	0	0	0	0	0	0
a33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6496	0
a34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6811
a35	0	0	0	0	0	0	0	0	0,6582	0	0	0	0	0	0	0
a36	0	0	0	0	0	0	0	0	0	0,6117	0	0	0	0	0	0
a37	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6497	0	0
a38	0	0	0	0	0	0	0	0	0	0,6044	0	0	0	0	0	0
a39	0	0	0	0	0	0	0	0	0	0	0,6987	0	0	0	0	0
a40	0	0	0	0	0	0	0	0	0,6579	0	0	0	0	0	0	0

**Tabla 3.14.** Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,4$

- $\beta = 0,6$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0	0,417	0,4531	0	0	
a2	0	0	0	0	0	0	0	0	0	0,6465						
a3	0	0	0	0	0	0	0	0	0,6486	0						
a4	0	0	0	0	0	0	0	0	0	0,4347	0	0,4185	0			
a5	0	0	0	0	0	0	0	0	0	0	0	0,4235	0	0,4103		
a6	0	0	0	0	0	0	0	0	0	0	0	0,4252	0	0,4282		
a7	0	0	0	0	0	0	0	0,3962	0	0	0	0	0	0,4619		
a8	0	0	0	0	0	0	0	0	0	0	0	0	0,4949	0,4211		
a9	0	0	0	0	0	0	0	0	0	0	0	0,434	0	0,4318		
a10	0	0	0	0	0	0	0	0	0	0,5708						
a11	0	0	0	0	0	0	0	0	0	0,6212						
a12	0	0	0	0	0	0	0	0	0	0	0	0	0,4781	0,4607		
a13	0	0	0	0	0	0	0	0	0	0	0	0	0,422	0		
a14	0	0	0	0	0	0	0	0	0	0	0,6928					
a15	0	0	0	0	0	0	0	0,5161	0	0	0	0	0	0,4537		
a16	0	0	0	0	0	0	0	0	0	0	0,4231	0	0,5593			
a17	0	0	0	0	0	0	0	0	0	0	0,6123	0	0			
a18	0	0	0	0	0	0,5514	0	0	0	0,4482						
a19	0	0	0,4842	0	0	0	0	0	0	0,4496						
a20	0,4279	0	0	0	0	0	0	0	0	0,4302						
a21	0	0	0	0	0	0	0	0	0	0	0	0	0,5644	0		
a22	0	0	0	0	0	0	0	0	0,4297	0	0	0	0	0,3795	0	0
a23	0	0	0	0	0	0	0	0	0,5109	0	0					
a24	0	0	0	0	0	0,4146	0,4058	0	0	0	0	0	0	0		
a25	0	0,4555	0	0	0	0	0	0	0	0	0	0	0	0,4429		
a26	0	0	0	0	0	0	0	0	0	0,7213						
a27	0	0	0	0	0	0	0	0,518	0	0	0	0,4169	0	0		
a28	0	0	0	0	0	0	0	0	0	0,4304	0	0	0	0,5183		
a29	0	0	0	0,3971	0	0	0	0	0	0	0,4312	0				
a30	0	0	0	0	0	0	0	0,5646	0	0						
a31	0	0	0	0	0	0	0	0	0	0,4583	0	0	0	0,5286		
a32	0	0	0	0	0	0	0	0	0,4293	0	0	0,5292				
a33	0	0	0	0	0	0	0,4318	0	0	0	0	0	0	0	0,5257	
a34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,4291	0,5556
a35	0	0	0	0	0	0	0	0,4202	0,5295	0						
a36	0	0	0	0	0	0	0,4438	0	0	0,4488						
a37	0	0	0	0	0	0	0	0	0	0	0,4492	0	0	0,5263		
a38	0	0	0	0	0	0	0	0	0	0	0	0,6019	0	0		
a39	0	0	0	0	0	0	0	0	0	0	0,5803	0	0	0		
a40	0	0	0	0	0	0	0	0	0,5087							

**Tabla 3.15.** Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,6$



$\beta = 0,7$ 

Tipo	Apuestas totales	Ganancia
1	0,6973	0
2	0,5981	0
3	0,9534	0
4	0,6723	0
5	0,6525	0
6	0,6698	0
7	0,6777	0
8	0,7571	0
9	0,6916	17,644
10	0,9123	0
11	0,566	0
12	0,7799	0
13	0,9184	0
14	0,6541	0
15	0,8159	0
16	0,8297	0
17	0,913	0
18	0,8307	0
19	0,7651	0
20	0,8138	0
21	0,8508	0
22	0,9539	0
23	0,8635	0
24	0,7626	0
25	0,7306	0
26	0,6849	0
27	0,7689	0
28	0,7808	0
29	0,6441	0
30	0,8881	0
31	0,8314	0
32	0,7931	0
33	0,8059	0
34	0,8298	0
35	0,779	0
36	0,7139	0
37	0,8302	0
38	0,903	0
39	0,8489	0
40	0,8923	0
Total	31,3244	17,644

 $\beta = 0,8$ 

Tipo	Apuestas totales	Ganancia
1	0,5246	0
2	0,8961	0
3	0,8096	0
4	0,8238	0
5	0,721	0
6	0,7824	0
7	0,7968	5,9533
8	0,8318	0
9	0,7602	13,5068
10	0,7601	0
11	0,9109	0
12	0,9163	0
13	0,7778	0
14	0,9026	0
15	0,9112	0
16	0,7696	0
17	0,766	0
18	0,6618	0
19	0,8592	0
20	0,9356	0
21	0,9373	0
22	0,6712	0
23	0,9721	0
24	0,8613	0
25	0,8243	0
26	0,8866	0
27	0,8827	0
28	0,8369	8,8607
29	0,7479	0
30	0,7423	0
31	0,9232	0
32	0,6278	0
33	0,9222	0
34	0,9199	0
35	0,8616	0
36	0,8895	0
37	0,6848	0
38	0,7553	0
39	0,9548	0
40	0,741	0
Total	32,7601	28,3208

**Tabla 3.16.** Validación cruzada: apuestas totales y ganancias para cada tipo, con  $\beta = 0,7$  y  $\beta = 0,8$

- $\beta = 0,7$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0	0	0,328	0,3693	0	0
a2	0	0	0	0	0	0	0	0	0	0,5981						
a3	0	0	0	0	0	0	0	0	0,6101	0,3433						
a4	0	0	0	0	0	0	0	0	0	0,346	0	0,3263	0			
a5	0	0	0	0	0	0	0	0	0	0	0	0,3332	0	0,3193		
a6	0	0	0	0	0	0	0	0	0	0	0	0,3346	0	0,3352		
a7	0	0	0	0	0	0	0	0,2987	0	0	0	0	0	0,379		
a8	0	0	0	0	0	0	0	0	0	0	0	0	0,4216	0,3355		
a9	0	0	0	0	0	0	0	0	0	0	0	0,3456	0	0,346		
a10	0	0	0	0	0	0	0	0	0,3824	0,5299						
a11	0	0	0	0	0	0	0	0	0	0,566						
a12	0	0	0	0	0	0	0	0	0	0	0	0	0,4001	0,3798		
a13	0	0	0	0	0	0	0	0	0	0	0	0	0,3419	0,5765		
a14	0	0	0	0	0	0	0	0	0	0	0,6541					
a15	0	0	0	0	0	0	0	0,4443	0	0	0	0	0	0	0,3716	
a16	0	0	0	0	0	0	0	0	0	0	0,3347	0	0,495			
a17	0	0	0	0	0	0	0	0	0	0	0,5642	0,3488	0			
a18	0	0	0	0	0	0,475	0	0	0	0,3557						
a19	0	0	0,4019	0	0	0	0	0	0	0,3632						
a20	0,3404	0	0	0	0	0	0,4734	0	0	0						
a21	0	0	0	0	0	0	0	0	0	0	0	0	0,5075	0,3433		
a22	0	0	0	0	0	0	0	0	0,35	0	0	0,3125	0	0,2914	0	0
a23	0	0	0	0	0	0	0	0	0,4482	0,4153	0					
a24	0	0	0	0	0	0	0,3175	0	0	0	0	0	0,4451	0		
a25	0	0,3726	0	0	0	0	0	0	0	0	0	0	0	0,358		
a26	0	0	0	0	0	0	0	0	0	0,6849						
a27	0	0	0	0	0	0	0	0,4442	0	0	0	0,3247	0	0		
a28	0	0	0	0	0	0	0	0	0	0,3395	0	0	0	0,4413		
a29	0	0	0	0,2999	0	0	0	0	0	0	0,3442	0				
a30	0	0	0	0	0	0	0	0,5092	0	0,3789						
a31	0	0	0	0	0	0	0	0	0	0,3754	0	0	0	0,456		
a32	0	0	0	0	0	0	0	0	0,3395	0	0	0,4536				
a33	0	0	0	0	0	0	0,3482	0	0	0	0	0	0	0	0,4577	
a34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,34	0,4898
a35	0	0	0	0	0	0	0	0,3244	0,4546	0						
a36	0	0	0	0	0	0	0,3529	0	0	0,361						
a37	0	0	0	0	0	0	0	0	0	0	0,3701	0	0	0,4601		
a38	0	0	0	0	0	0	0	0	0	0,3523	0	0,5507	0	0		
a39	0	0	0	0	0	0	0	0	0	0	0,5275	0	0,3214	0		
a40	0	0	0	0	0	0	0	0,4516	0,4407							

**Tabla 3.17.** Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,7$

- $\beta = 0,8$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a <sub>1</sub>	0	0	0	0	0	0	0	0	0	0	0	0,239	0,2856	0	0	
a <sub>2</sub>	0	0	0	0	0	0	0	0	0,3413	0,5548						
a <sub>3</sub>	0	0	0	0	0	0	0	0	0,5585	0,2511						
a <sub>4</sub>	0	0	0,3146	0	0	0	0	0	0	0,2654	0	0,2438	0			
a <sub>5</sub>	0	0	0	0	0	0	0	0	0	0	0	0,249	0,24	0,232		
a <sub>6</sub>	0	0	0	0	0	0	0	0	0	0,2789	0	0,2499	0	0,2536		
a <sub>7</sub>	0	0	0	0	0	0,2835	0	0,2127	0	0	0	0	0	0,3006		
a <sub>8</sub>	0	0	0	0	0	0	0	0	0	0	0	0,2266	0,3518	0,2534		
a <sub>9</sub>	0	0	0	0	0	0	0	0	0	0	0	0,2665	0,2289	0,2648		
a <sub>10</sub>	0	0	0	0	0	0	0	0	0,2957	0,4644						
a <sub>11</sub>	0	0	0	0	0	0	0	0	0,3841	0,5268						
a <sub>12</sub>	0	0	0	0	0	0	0	0	0,2867	0	0	0	0,3264	0,3032		
a <sub>13</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,2555	0,5223		
a <sub>14</sub>	0	0	0	0	0	0	0	0	0,279	0	0,6236					
a <sub>15</sub>	0	0	0	0	0	0	0	0,3799	0,2345	0	0	0	0	0,2968		
a <sub>16</sub>	0	0	0	0	0	0	0	0	0	0	0	0,338	0,4316			
a <sub>17</sub>	0	0	0	0	0	0	0	0	0	0	0,5074	0,2586	0			
a <sub>18</sub>	0	0	0	0	0	0,3986	0	0	0	0,2632						
a <sub>19</sub>	0	0	0,3328	0	0	0	0,2392	0	0	0,2872						
a <sub>20</sub>	0,2571	0	0	0	0	0	0,4153	0	0	0,2632						
a <sub>21</sub>	0	0	0	0	0	0	0	0	0,2293	0	0	0	0,4476	0,2604		
a <sub>22</sub>	0	0	0	0	0	0	0	0	0,261	0	0	0,2161	0	0,1941	0	0
a <sub>23</sub>	0	0	0	0	0	0	0	0	0,3808	0,3444	0,2469					
a <sub>24</sub>	0	0	0	0	0	0,246	0,2338	0	0	0	0	0	0,3815	0		
a <sub>25</sub>	0	0,2949	0	0	0	0	0	0	0	0	0	0,2512	0	0,2782		
a <sub>26</sub>	0	0	0	0	0	0	0,2341	0	0	0,6525						
a <sub>27</sub>	0	0	0	0	0	0	0	0,3746	0	0	0	0,2394	0,2687	0		
a <sub>28</sub>	0	0	0	0	0	0	0	0	0,2161	0,252	0	0	0	0,3688		
a <sub>29</sub>	0	0	0	0,2165	0,2687	0	0	0	0	0	0,2627	0				
a <sub>30</sub>	0	0	0	0	0	0	0	0,4459	0	0,2964						
a <sub>31</sub>	0	0	0	0	0	0	0	0	0,2375	0,2962	0	0	0	0,3895		
a <sub>32</sub>	0	0	0	0	0	0	0	0	0,2498	0	0	0,378				
a <sub>33</sub>	0	0	0	0	0	0	0,2699	0	0,2573	0	0	0	0	0	0,395	
a <sub>34</sub>	0	0	0	0	0	0	0	0,2366	0	0	0	0	0	0	0,2569	0,4264
a <sub>35</sub>	0	0	0	0	0	0	0,2322	0,2413	0,3881	0						
a <sub>36</sub>	0	0	0	0,3354	0	0	0,2739	0	0	0,2802						
a <sub>37</sub>	0	0	0	0	0	0	0	0	0	0	0,291	0	0	0,3938		
a <sub>38</sub>	0	0	0	0	0	0	0	0	0	0,2646	0	0,4907	0	0		
a <sub>39</sub>	0	0	0	0	0	0	0	0	0	0	0,4716	0,2473	0,2359	0		
a <sub>40</sub>	0	0	0	0	0	0	0	0,376	0,365							

**Tabla 3.18.** Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,8$

$\beta = 0,9$ 

Tipo	Apuestas totales	Ganancia
1	0,9636	0
2	0,9118	0
3	0,8525	0
4	0,885	3,6635
5	0,8853	0
6	0,8701	0
7	0,9526	4,3978
8	0,9354	0
9	0,8551	9,3958
10	0,9851	0
11	0,7701	0
12	0,9951	0
13	0,9537	0
14	0,9415	0
15	0,9368	0
16	0,9551	0
17	0,8744	0
18	0,9384	0
19	0,976	0
20	0,903	0
21	0,9791	0
22	0,9962	0
23	0,9039	0
24	0,9601	0
25	0,9003	0
26	0,936	0
27	0,8588	0
28	0,9219	5,147
29	0,9785	0
30	0,8058	0
31	0,8842	0
32	0,931	0
33	0,8846	0
34	0,9631	0
35	0,7729	0
36	0,9814	0
37	0,8256	14,1576
38	0,8722	0
39	0,9258	0
40	0,9857	0
<b>Total</b>	<b>36,6077</b>	<b>36,7617</b>

 $\beta = 1$ 

Tipo	Apuestas totales	Ganancia
1	0,9101	0
2	0,8017	0
3	0,9535	0,2032
4	0,9053	1,3762
5	0,7429	0,3806
6	0,5871	0,3961
7	0,6501	2,3841
8	0,631	0,1452
9	0,5683	4,9031
10	0,7563	0,1151
11	0,875	1,4325
12	0,8808	0
13	0,9371	0
14	0,7631	0,027
15	0,5314	0,1407
16	0,9205	0,7076
17	0,7335	0
18	0,9325	0,8487
19	0,9416	0,7856
20	0,8743	0
21	0,9212	0,0923
22	0,5194	0
23	0,9456	0
24	0,7933	0
25	0,9084	0
26	0,8727	1,1814
27	0,861	0,1172
28	0,8543	1,2807
29	0,8242	0,5317
30	0,952	0,0403
31	0,9834	0
32	0,7913	0
33	0,8663	0,192
34	0,642	0
35	0,8309	0,2494
36	0,7087	0,8842
37	0,9716	10,7324
38	0,9372	0
39	0,9863	0,7276
40	0,8638	0
<b>Total</b>	<b>32,9297</b>	<b>29,8749</b>

**Tabla 3.19.** Validación cruzada: apuestas totales y ganancias para cada tipo, con  $\beta = 0,9$  y  $\beta = 1$

- $\beta = 0,9$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0,4498	0,1561	0,211	0	0,1467	
a2	0	0	0	0	0	0	0,1368	0	0,2676	0,5074						
a3	0	0	0	0,1725	0	0	0	0	0,5123	0,1677						
a4	0	0	0,2399	0	0	0	0	0,1629	0	0,1826	0,1409	0,1587	0			
a5	0	0	0	0	0	0	0,1943	0	0	0,2197	0	0,1672	0,158	0,1461		
a6	0	0,1674	0	0	0,1652	0	0	0	0	0,2018	0	0,1647	0	0,171		
a7	0	0	0	0	0	0,2094	0,2196	0,1281	0	0	0,1727	0	0,2228			
a8	0	0	0,138	0	0	0	0	0	0	0	0,1998	0,1434	0,2824	0,1718		
a9	0	0	0	0	0,1455	0	0	0	0	0	0,1908	0,1885	0,1461	0,1842		
a10	0	0	0,1763	0	0	0	0	0,1632	0,228	0,4176						
a11	0	0	0	0	0	0	0	0	0,302	0,4681						
a12	0	0	0	0	0	0	0,1438	0	0,208	0	0,1656	0	0,2519	0,2258		
a13	0	0	0	0	0	0	0,1317	0	0	0	0,1754	0	0,1722	0,4744		
a14	0	0	0	0	0,1537	0	0	0	0,2006	0	0,5872					
a15	0	0	0,1298	0	0	0	0	0,3135	0,1502	0,1234	0	0	0	0,2199		
a16	0	0	0	0	0	0	0	0	0,1496	0	0,1663	0,2685	0,3707			
a17	0	0	0	0	0	0	0	0,1286	0	0	0,4555	0,1783	0,112			
a18	0	0	0	0	0	0,3445	0,18	0,2239	0	0,19						
a19	0	0	0,2647	0	0	0,1426	0,1603	0,1962	0	0,2122						
a20	0,1699	0	0	0	0	0	0,3468	0,2101	0	0,1762						
a21	0	0	0	0	0	0	0	0	0,1417	0	0	0,2737	0,387	0,1767		
a22	0	0	0	0	0	0,1155	0	0	0,1852	0,257	0	0,1375	0	0,1099	0	0,1911
a23	0	0	0	0	0	0	0	0,169	0,3099	0,2683	0,1567					
a24	0	0	0	0	0	0,1644	0,1501	0	0,1812	0	0	0,1465	0,3179	0		
a25	0	0,2168	0	0	0	0	0	0,1751	0	0,1419	0	0,1685	0	0,198		
a26	0	0	0	0	0	0	0,148	0	0,1704	0,6176						
a27	0	0	0	0	0	0	0	0,3029	0	0,2216	0	0,1507	0,1836	0		
a28	0	0	0	0	0	0	0,1287	0	0,1255	0,1649	0,206	0	0	0,2968		
a29	0	0	0	0,1391	0,2008	0	0	0,1453	0	0,1564	0,1838	0,1531				
a30	0	0	0	0	0	0,2012	0	0,3862	0	0,2184						
a31	0	0	0	0	0	0	0	0	0,1488	0,215	0,2007	0	0	0,3197		
a32	0	0	0	0	0	0	0	0	0,1666	0,1592	0,2889	0,3163				
a33	0	0	0	0	0	0	0,1885	0	0,1742	0	0	0	0	0,1926	0,3293	
a34	0	0	0	0	0	0	0	0,151	0	0	0,1338	0,1432	0	0	0,1726	0,3625
a35	0	0	0	0	0	0	0,1389	0,1495	0,316	0,1685						
a36	0	0	0	0,2626	0	0,1472	0,1921	0,1817	0	0,1978						
a37	0	0	0	0	0	0	0	0	0	0,2776	0,2162	0	0	0,3318		
a38	0	0	0	0	0	0	0	0	0	0,1809	0,2554	0,4359	0	0		
a39	0	0	0	0	0	0	0	0	0	0	0,413	0,1606	0,1479	0,2043		
a40	0	0	0,2477	0	0	0,1327	0	0,3103	0,295							

**Tabla 3.20.** Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,9$

- $\beta = 1$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a <sub>1</sub>	0,0074	0	0	0,0581	0,0087	0,004	0	0,0332	0,0923	0,0215	0,3937	0,0683	0,129	0,0365	0,0574	
a <sub>2</sub>	0	0	0,0223	0	0	0,0436	0,043	0,0445	0,1902	0,4581						
a <sub>3</sub>	0	0,0369	0,0234	0,0938	0,0206	0,058	0,1492	0,0166	0,4679	0,0871						
a <sub>4</sub>	0	0	0,1632	0,1088	0,065	0,0313	0,0623	0,0776	0,0584	0,0988	0,0529	0,0724	0,1146			
a <sub>5</sub>	0	0	0,0586	0,0346	0	0,0237	0,105	0,0603	0,1259	0,1344	0	0,0778	0,067	0,0556		
a <sub>6</sub>	0,0609	0,0626	0	0	0,0602	0,0273	0,0583	0,0167	0	0,1029	0,0443	0,0722	0,0076	0,0741		
a <sub>7</sub>	0	0	0	0,0853	0	0,1135	0,1215	0,0259	0,0604	0	0,0671	0,0383	0,1381			
a <sub>8</sub>	0	0	0,0368	0,0586	0,0132	0,01	0,0413	0,0043	0	0,0222	0,1055	0,0482	0,2069	0,084		
a <sub>9</sub>	0	0	0	0,0181	0,0455	0	0,0009	0,0296	0,1378	0	0,0958	0,0958	0,0487	0,0961		
a <sub>10</sub>	0	0,0144	0,0801	0,0016	0	0,0214	0,0876	0,0655	0,1375	0,3482						
a <sub>11</sub>	0	0,0065	0,0146	0,0331	0	0	0,0309	0,1302	0,2404	0,4193						
a <sub>12</sub>	0	0	0	0,0051	0,0093	0,0253	0,0527	0,0728	0,1247	0,0527	0,0776	0,1406	0,1745	0,1455		
a <sub>13</sub>	0	0	0,0241	0,0097	0,0407	0,0053	0,0418	0,0258	0,0377	0,0029	0,0903	0,1491	0,0871	0,4226		
a <sub>14</sub>	0	0	0	0	0,0555	0	0	0,0492	0,1112	0,0027	0,5445					
a <sub>15</sub>	0	0	0,023	0	0	0	0,0297	0,2343	0,0493	0,0196	0	0,0235	0,0216	0,1304		
a <sub>16</sub>	0	0	0,0113	0,0323	0,0115	0,0434	0,0407	0,0914	0,0613	0,0472	0,0802	0,1936	0,3076			
a <sub>17</sub>	0,009	0	0	0,0256	0,0215	0	0,1207	0,0319	0,0241	0	0,3988	0,0883	0,0136			
a <sub>18</sub>	0	0,0849	0,1096	0,0316	0,0681	0,2763	0,0935	0,1423	0,0216	0,1046						
a <sub>19</sub>	0,2293	0,0982	0,1892	0	0,0361	0,053	0,0729	0,1123	0,0195	0,1311						
a <sub>20</sub>	0,0828	0	0,0615	0,0785	0,0196	0,1034	0,2787	0,1265	0,0339	0,0894						
a <sub>21</sub>	0	0	0,0251	0,0773	0,0019	0,0262	0,0084	0,0523	0,0521	0,0172	0,047	0,1977	0,325	0,091		
a <sub>22</sub>	0	0,0166	0	0,0007	0,0688	0,0066	0	0	0,092	0,1541	0,0424	0,0353	0	0,0084	0,0154	0,0791
a <sub>23</sub>	0,0166	0,0701	0,059	0	0,0314	0,0914	0,0805	0,087	0,2411	0,1955	0,073					
a <sub>24</sub>	0	0,0036	0,0682	0,0433	0,0289	0,0737	0,0586	0,0277	0,0884	0,0213	0,0278	0,0523	0,2428	0,0567		
a <sub>25</sub>	0,0534	0,1374	0,0485	0,0608	0	0	0,0534	0,0918	0,1221	0,0545	0,0332	0,0839	0,0529	0,1165		
a <sub>26</sub>	0	0	0,0125	0,0909	0,0113	0,0065	0,058	0,0302	0,0825	0,5808						
a <sub>27</sub>	0	0	0,0227	0,0774	0,0156	0,0196	0,019	0,2316	0	0,1417	0,0744	0,0627	0,0993	0,097		
a <sub>28</sub>	0	0,023	0,0269	0,0816	0,0086	0,0216	0,0345	0,0367	0,0312	0,0756	0,1198	0,0414	0,1316	0,2218		
a <sub>29</sub>	0,1436	0	0,1901	0,0434	0,1102	0,0072	0,0709	0,0459	0	0,0597	0,0972	0,056				
a <sub>30</sub>	0	0,0309	0,0205	0,0058	0,108	0,125	0,0883	0,3278	0,1033	0,1424						
a <sub>31</sub>	0,0327	0	0,0525	0,0109	0,0211	0,128	0,0337	0,0076	0,0634	0,1359	0,1216	0,0223	0,1004	0,2533		
a <sub>32</sub>	0,0571	0	0	0,0173	0	0,008	0,0366	0,0769	0,0772	0,0657	0,211	0,2415				
a <sub>33</sub>	0	0	0,0128	0,0195	0,0077	0,0816	0,1075	0,0271	0,0916	0,0123	0,0771	0,0308	0,0224	0,1119	0,264	
a <sub>34</sub>	0	0,0104	0	0,0259	0	0	0	0,0481	0,0142	0,0728	0,0289	0,0394	0,0105	0,0173	0,0795	0,295
a <sub>35</sub>	0	0,0036	0,3308	0	0	0,0416	0,0538	0,0652	0,2487	0,0872						
a <sub>36</sub>	0	0,0054	0	0,1731	0	0,0437	0,0981	0,0839	0,1965	0,108						
a <sub>37</sub>	0	0	0,0173	0,0096	0,0017	0,0024	0,0174	0,0682	0,0355	0,2104	0,1414	0,1733	0,0245	0,2699		
a <sub>38</sub>	0	0	0	0,0325	0,0067	0,0719	0,0073	0,0118	0,0262	0,0963	0,1795	0,38	0,0478	0,0772		
a <sub>39</sub>	0,0133	0	0,1039	0,0422	0,0221	0,0407	0,0015	0,1007	0,0079	0,0353	0,3561	0,076	0,0616	0,125		
a <sub>40</sub>	0,107	0	0,1629	0,0856	0,0036	0,0351	0,0143	0,2355	0,2198							

**Tabla 3.21.** Validación cruzada: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 1$

Como vemos, la diferencia más grande entre lo que apostamos y lo que ganamos se da con  $\beta = 0$ . Con ese valor tenemos un beneficio de 9,7779€. Por lo tanto, en este caso podríamos decir que el valor de  $\beta$  óptimo sería 0.

Sin embargo, observando lo que apostamos a cada caballo en cada carrera, podemos decir que hasta  $\beta = 0,6$  solemos apostar únicamente por un caballo por carrera, y, además, con  $\beta = 0$  hasta  $\beta = 0,7$  sólo estamos obteniendo ganancias para una carrera, aunque el beneficio sea muy grande. Esto lleva un gran riesgo, puesto que si el caballo por el que hemos apostado no es el ganador, situación que se puede dar con mucha probabilidad, tendremos una gran pérdida.

Pongamos un ejemplo, suponiendo equiprobabilidad en que gane cada caballo, y teniendo en cuenta que es una carrera con diez caballos. La probabilidad de que gane cada caballo será de  $\frac{1}{10} = 0,1$  y, por lo tanto, la probabilidad de que obtengamos beneficios apostando sólo por un caballo será de un 10%, frente al 90% de que perdamos dinero. Por lo tanto, podemos concluir que es muy arriesgado apostar sólo por un caballo en cada carrera.

Si nos fijamos en los valores de apuestas para cada caballo con  $\beta = 0,9$  y  $\beta = 1$  vemos que el número de caballos por los que se apuesta en cada carrera aumenta de forma considerable.

Además con  $\beta = 0,9$  obtenemos un beneficio de 0,154 €. Si lo comparamos con los 9,7 € de  $\beta = 0$  vemos que es un valor muy bajo. En cambio, en este caso tenemos más posibilidades de acertar el ganador de una carrera, ya que estamos apostando por un mayor número de caballos.

Si seguimos con el ejemplo anterior, suponiendo que apostamos por cuatro caballos. La probabilidad de que ganemos será de un 40% y de que no acertemos de un 60%. Si lo comparamos con el 10%, vemos que, en este caso, nuestra apuesta lleva un menor nivel de riesgo.

Por lo tanto, bajo nuestro punto de vista, sería mejor emplear un valor de  $\beta$  de 0,9, ya que tenemos beneficios, y además nos aseguramos de que no vamos a tener un gran número de pérdidas si no acertamos en nuestra predicción. Así, nuestro valor óptimo de  $\beta$  será de 0,9.

En cambio, los apostantes arriesgados (sección 1.5.), podrían decidir que el mejor valor para apostar es de 0.

Como conclusión podemos decir que, dependiendo del tipo de apostante que seamos, elegiremos un valor de  $\beta$  u otro.

### 3.1.3.2. Estudio LOO

$\beta = 0$ 

Tipo	Apuestas totales	Ganancia
1	0,9414	0
2	0,9368	0
3	0,9418	0
4	0,9427	0
5	0,942	0
6	0,9486	0
7	0,9442	0
8	0,9204	0
9	0,9279	47,3216
10	0,9369	0
11	0,9527	0
12	0,9322	0
13	0,9373	0
14	0,925	0
15	0,9178	0
16	0,9353	0
17	0,9427	0
18	0,9573	0
19	0,9276	0
20	0,9484	0
21	0,937	0
22	0,9283	0
23	0,9367	0
24	0,9286	0
25	0,9335	0
26	0,9393	0
27	0,9462	0
28	0,9607	0
29	0,9342	0
30	0,9356	0
31	0,9409	0
32	0,9539	0
33	0,9146	0
34	0,9409	0
35	0,9446	0
36	0,9557	0
37	0,9099	0
38	0,9515	0
39	0,9354	0
40	0,9562	0
Total	37,5427	47,3216

 $\beta = 0,05$ 

Tipo	Apuestas totales	Ganancia
1	0,8992	0
2	0,9116	0
3	0,9099	0
4	0,9007	0
5	0,9013	0
6	0,9053	0
7	0,9082	0
8	0,8827	0
9	0,888	45,2868
10	0,9001	0
11	0,9316	0
12	0,8955	0
13	0,8956	0
14	0,8986	0
15	0,8794	0
16	0,9014	0
17	0,9148	0
18	0,9129	0
19	0,8897	0
20	0,9028	0
21	0,905	0
22	0,8871	0
23	0,9067	0
24	0,8832	0
25	0,8926	0
26	0,9259	0
27	0,9059	0
28	0,917	0
29	0,8919	0
30	0,907	0
31	0,8998	0
32	0,9089	0
33	0,8845	0
34	0,9051	0
35	0,916	0
36	0,9137	0
37	0,8738	0
38	0,9095	0
39	0,901	0
40	0,924	0
Total	36,0879	45,2868

Tabla 3.22. Leave-One-Out: Apuestas totales y ganancias para cada tipo, con  $\beta=0$  y  $\beta=0,05$



- $\beta = 0$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0	0,9414	0	0	0	
a2	0	0	0	0	0	0	0	0	0	0,9368						
a3	0	0	0	0	0	0	0	0	0,9418	0						
a4	0	0	0	0	0	0	0	0	0	0,9427	0	0	0			
a5	0	0	0	0	0	0	0	0	0	0	0	0	0	0,942		
a6	0	0	0	0	0	0	0	0	0	0	0	0,9486	0	0		
a7	0	0	0	0	0	0	0	0	0	0	0	0	0,9442			
a8	0	0	0	0	0	0	0	0	0	0	0	0	0,9204	0		
a9	0	0	0	0	0	0	0	0	0	0	0	0	0	0,9279		
a10	0	0	0,9369	0	0	0	0	0	0	0						
a11	0	0	0	0	0	0	0	0	0	0,9527						
a12	0	0	0	0	0	0	0	0	0	0	0	0	0,9322	0		
a13	0	0	0	0	0	0	0	0	0	0	0	0	0,9373	0		
a14	0	0	0	0	0	0	0	0	0	0	0,925					
a15	0	0	0	0	0	0	0	0,9178	0	0	0	0	0	0		
a16	0	0	0	0	0	0	0	0	0	0	0	0	0,9353			
a17	0	0	0	0	0	0	0	0	0	0	0,9427	0	0			
a18	0	0	0	0	0	0	0	0	0	0,9573						
a19	0	0	0	0	0	0	0	0	0	0,9276						
a20	0,9484	0	0	0	0	0	0	0	0	0						
a21	0	0	0	0	0	0	0	0	0	0	0	0	0,937	0		
a22	0	0	0	0	0	0	0	0	0,9293	0	0	0	0	0	0	0
a23	0	0	0	0	0	0	0	0	0,9367	0	0					
a24	0	0	0	0	0	0	0,9286	0	0	0	0	0	0	0		
a25	0	0,9335	0	0	0	0	0	0	0	0	0	0	0	0		
a26	0	0	0	0	0	0	0	0	0	0,9393						
a27	0	0	0	0	0	0	0	0,9462	0	0	0	0	0	0		
a28	0	0	0	0	0	0	0	0	0	0,9607	0	0	0	0		
a29	0	0	0	0	0	0	0	0	0	0	0,9342	0				
a30	0	0	0	0	0	0	0	0,9356	0	0						
a31	0	0	0	0	0	0	0	0	0	0,9409	0	0	0	0		
a32	0	0	0	0	0	0	0	0	0,9539	0	0	0				
a33	0	0	0	0	0	0	0,9146	0	0	0	0	0	0	0	0	
a34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,9409
a35	0	0	0	0	0	0	0	0	0,9446	0						
a36	0	0	0	0	0	0	0	0	0	0,9557						
a37	0	0	0	0	0	0	0	0	0,9099	0	0	0	0	0		
a38	0	0	0	0	0	0	0	0	0	0,9515	0	0	0	0		
a39	0	0	0	0	0	0	0	0	0	0	0,9354	0	0	0		
a40	0	0	0	0	0	0	0	0	0,9562							

**Tabla 3.23.** *Leave-One-Out*: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0$

- $\beta = 0,05$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0	0,8992	0	0	0	
a2	0	0	0	0	0	0	0	0	0	0,9116						
a3	0	0	0	0	0	0	0	0	0,9099	0						
a4	0	0	0	0	0	0	0	0	0	0,9007	0	0	0			
a5	0	0	0	0	0	0	0	0	0	0	0	0	0	0,9013		
a6	0	0	0	0	0	0	0	0	0	0	0	0,9053	0	0		
a7	0	0	0	0	0	0	0	0	0	0	0	0	0,9082			
a8	0	0	0	0	0	0	0	0	0	0	0	0	0,8827	0		
a9	0	0	0	0	0	0	0	0	0	0	0	0	0	0,888		
a10	0	0	0	0	0	0	0	0	0	0,9001						
a11	0	0	0	0	0	0	0	0	0	0,9316						
a12	0	0	0	0	0	0	0	0	0	0	0	0	0,8955	0		
a13	0	0	0	0	0	0	0	0	0	0	0	0	0,8956	0		
a14	0	0	0	0	0	0	0	0	0	0	0,8986					
a15	0	0	0	0	0	0	0	0,8794	0	0	0	0	0	0	0	
a16	0	0	0	0	0	0	0	0	0	0	0	0	0,9014			
a17	0	0	0	0	0	0	0	0	0	0	0,9148	0	0			
a18	0	0	0	0	0	0	0	0	0	0,9129						
a19	0	0	0	0	0	0	0	0	0	0,8897						
a20	0,9028	0	0	0	0	0	0	0	0	0						
a21	0	0	0	0	0	0	0	0	0	0	0	0	0,905	0		
a22	0	0	0	0	0	0	0	0	0,8871	0	0	0	0	0	0	0
a23	0	0	0	0	0	0	0	0	0,9067	0	0					
a24	0	0	0	0	0	0	0,8832	0	0	0	0	0	0	0	0	
a25	0	0,8926	0	0	0	0	0	0	0	0	0	0	0	0	0	
a26	0	0	0	0	0	0	0	0	0	0,9259						
a27	0	0	0	0	0	0	0	0,9059	0	0	0	0	0	0	0	
a28	0	0	0	0	0	0	0	0	0	0,917	0	0	0	0	0	
a29	0	0	0	0	0	0	0	0	0	0	0,8919	0				
a30	0	0	0	0	0	0	0	0,907	0	0						
a31	0	0	0	0	0	0	0	0	0	0,8998	0	0	0	0		
a32	0	0	0	0	0	0	0	0	0,9089	0	0	0				
a33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8845	
a34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,9051
a35	0	0	0	0	0	0	0	0	0,916	0						
a36	0	0	0	0	0	0	0	0	0	0,9137						
a37	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8738		
a38	0	0	0	0	0	0	0	0	0	0,9095	0	0	0	0		
a39	0	0	0	0	0	0	0	0	0	0	0,901	0	0	0		
a40	0	0	0	0	0	0	0	0	0,924							

**Tabla 3.24.** *Leave-One-Out*: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,05$

$\beta = 0,1$ 

Tipo	Apuestas totales	Ganancia
1	0,8569	0
2	0,8863	0
3	0,878	0
4	0,8587	0
5	0,8606	0
6	0,8621	0
7	0,8722	0
8	0,8451	0
9	0,8481	43,252
10	0,8632	0
11	0,9106	0
12	0,8588	0
13	0,8538	0
14	0,8723	0
15	0,841	0
16	0,8676	0
17	0,8869	0
18	0,8685	0
19	0,8518	0
20	0,8573	0
21	0,8731	0
22	0,8449	0
23	0,8768	0
24	0,8378	0
25	0,8517	0
26	0,9125	0
27	0,8657	0
28	0,8734	0
29	0,8495	0
30	0,8784	0
31	0,8587	0
32	0,8639	0
33	0,8544	0
34	0,8692	0
35	0,8873	0
36	0,8716	0
37	0,8377	0
38	0,8674	0
39	0,8665	0
40	0,8918	0
<b>Total</b>	<b>34,6321</b>	<b>43,252</b>

 $\beta = 0,2$ 

Tipo	Apuestas totales	Ganancia
1	0,7724	0
2	0,8357	0
3	0,8143	0
4	0,7747	0
5	0,7792	0
6	0,7756	0
7	0,8001	0
8	0,7697	0
9	0,7683	39,1825
10	0,7895	0
11	0,8686	0
12	0,7854	0
13	0,7704	0
14	0,8196	0
15	0,7642	0
16	0,7998	0
17	0,8312	0
18	0,7797	0
19	0,7759	0
20	0,7662	0
21	0,8091	0
22	0,7604	0
23	0,817	0
24	0,7471	0
25	0,7698	0
26	0,8857	0
27	0,7851	0
28	0,786	0
29	0,7649	0
30	0,8212	0
31	0,7764	0
32	0,7738	0
33	0,7942	0
34	0,7976	0
35	0,8301	0
36	0,7875	0
37	0,7656	0
38	0,7833	0
39	0,7977	0
40	0,8273	0
<b>Total</b>	<b>31,7203</b>	<b>39,1825</b>

**Tabla 3.25.** *Leave-One-Out*: apuestas totales y ganancias para cada tipo, con  $\beta = 0,1$  y  $\beta = 0,2$

- $\beta = 0,1$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a <sub>1</sub>	0	0	0	0	0	0	0	0	0	0	0	0,8569	0	0	0	
a <sub>2</sub>	0	0	0	0	0	0	0	0	0	0,8863						
a <sub>3</sub>	0	0	0	0	0	0	0	0	0,878	0						
a <sub>4</sub>	0	0	0	0	0	0	0	0	0	0,8587	0	0	0			
a <sub>5</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8606		
a <sub>6</sub>	0	0	0	0	0	0	0	0	0	0	0	0,8621	0	0		
a <sub>7</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,8722			
a <sub>8</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,8451	0		
a <sub>9</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8481		
a <sub>10</sub>	0	0	0	0	0	0	0	0	0	0,8632						
a <sub>11</sub>	0	0	0	0	0	0	0	0	0	0,9106						
a <sub>12</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,8588	0		
a <sub>13</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,8538	0		
a <sub>14</sub>	0	0	0	0	0	0	0	0	0	0	0,8723					
a <sub>15</sub>	0	0	0	0	0	0	0	0,841	0	0	0	0	0	0	0	
a <sub>16</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,8676			
a <sub>17</sub>	0	0	0	0	0	0	0	0	0	0	0,8869	0	0			
a <sub>18</sub>	0	0	0	0	0	0	0	0	0	0,8685						
a <sub>19</sub>	0	0	0	0	0	0	0	0	0	0,8518						
a <sub>20</sub>	0,8573	0	0	0	0	0	0	0	0	0	0					
a <sub>21</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,8731	0		
a <sub>22</sub>	0	0	0	0	0	0	0	0	0,8449	0	0	0	0	0	0	0
a <sub>23</sub>	0	0	0	0	0	0	0	0	0,8768	0	0					
a <sub>24</sub>	0	0	0	0	0	0	0,8378	0	0	0	0	0	0	0		
a <sub>25</sub>	0	0,8517	0	0	0	0	0	0	0	0	0	0	0	0	0	
a <sub>26</sub>	0	0	0	0	0	0	0	0	0	0,9125						
a <sub>27</sub>	0	0	0	0	0	0	0	0,8657	0	0	0	0	0	0	0	
a <sub>28</sub>	0	0	0	0	0	0	0	0	0	0,8734	0	0	0	0	0	
a <sub>29</sub>	0	0	0	0	0	0	0	0	0	0	0,8495	0				
a <sub>30</sub>	0	0	0	0	0	0	0	0,8784	0	0						
a <sub>31</sub>	0	0	0	0	0	0	0	0	0	0,8587	0	0	0	0		
a <sub>32</sub>	0	0	0	0	0	0	0	0	0,8639	0	0	0				
a <sub>33</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8544	
a <sub>34</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8692
a <sub>35</sub>	0	0	0	0	0	0	0	0	0,8873	0						
a <sub>36</sub>	0	0	0	0	0	0	0	0	0	0,8716						
a <sub>37</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8377		
a <sub>38</sub>	0	0	0	0	0	0	0	0	0	0,8674	0	0	0	0		
a <sub>39</sub>	0	0	0	0	0	0	0	0	0	0	0,8665	0	0	0		
a <sub>40</sub>	0	0	0	0	0	0	0	0	0,8918							

**Tabla 3.26.** *Leave-One-Out*: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,1$

- $\beta = 0,2$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a <sub>1</sub>	0	0	0	0	0	0	0	0	0	0	0	0,7724	0	0	0	0
a <sub>2</sub>	0	0	0	0	0	0	0	0	0	0,8357						
a <sub>3</sub>	0	0	0	0	0	0	0	0	0,8143	0						
a <sub>4</sub>	0	0	0	0	0	0	0	0	0	0,7747	0	0	0			
a <sub>5</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7792		
a <sub>6</sub>	0	0	0	0	0	0	0	0	0	0	0	0,7756	0	0		
a <sub>7</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,8001			
a <sub>8</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,7697	0		
a <sub>9</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7683		
a <sub>10</sub>	0	0	0	0	0	0	0	0	0	0,7895						
a <sub>11</sub>	0	0	0	0	0	0	0	0	0	0,8686						
a <sub>12</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,7854	0		
a <sub>13</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,7704	0		
a <sub>14</sub>	0	0	0	0	0	0	0	0	0	0	0,8196					
a <sub>15</sub>	0	0	0	0	0	0	0	0,7642	0	0	0	0	0	0	0	
a <sub>16</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7998		
a <sub>17</sub>	0	0	0	0	0	0	0	0	0	0	0,8312	0	0			
a <sub>18</sub>	0	0	0	0	0	0	0	0	0	0,7797						
a <sub>19</sub>	0	0	0	0	0	0	0	0	0	0,7759						
a <sub>20</sub>	0,7662	0	0	0	0	0	0	0	0	0						
a <sub>21</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0,8091	0		
a <sub>22</sub>	0	0	0	0	0	0	0	0	0,7604	0	0	0	0	0	0	0
a <sub>23</sub>	0	0	0	0	0	0	0	0	0,817	0	0					
a <sub>24</sub>	0	0	0	0	0	0	0,7471	0	0	0	0	0	0	0	0	
a <sub>25</sub>	0	0,7698	0	0	0	0	0	0	0	0	0	0	0	0	0	
a <sub>26</sub>	0	0	0	0	0	0	0	0	0	0,8857						
a <sub>27</sub>	0	0	0	0	0	0	0	0,7851	0	0	0	0	0	0	0	
a <sub>28</sub>	0	0	0	0	0	0	0	0	0	0,786	0	0	0	0	0	
a <sub>29</sub>	0	0	0	0	0	0	0	0	0	0,7649	0					
a <sub>30</sub>	0	0	0	0	0	0	0	0,8212	0	0						
a <sub>31</sub>	0	0	0	0	0	0	0	0	0	0,7764	0	0	0	0		
a <sub>32</sub>	0	0	0	0	0	0	0	0	0,7738	0	0	0				
a <sub>33</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7942	
a <sub>34</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7976
a <sub>35</sub>	0	0	0	0	0	0	0	0	0,8301	0						
a <sub>36</sub>	0	0	0	0	0	0	0	0	0	0,7875						
a <sub>37</sub>	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7656		
a <sub>38</sub>	0	0	0	0	0	0	0	0	0	0,7833	0	0	0	0		
a <sub>39</sub>	0	0	0	0	0	0	0	0	0	0,7977	0	0	0			
a <sub>40</sub>	0	0	0	0	0	0	0	0	0,8273							

Tabla 3.27. *Leave-One-Out*: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,2$

$\beta = 0,4$ 

Tipo	Apuestas totales	Ganancia
1	0,6034	0
2	0,7346	0
3	0,6868	0
4	0,6066	0
5	0,6163	0
6	0,6026	0
7	0,656	0
8	0,6189	0
9	0,6087	31,0434
10	0,642	0
11	0,7845	0
12	0,6385	0
13	0,6035	0
14	0,7141	0
15	0,6105	0
16	0,6642	0
17	0,7198	0
18	0,6021	0
19	0,6242	0
20	0,854	0
21	0,6812	0
22	0,5916	0
23	0,6973	0
24	0,5655	0
25	0,606	0
26	0,8321	0
27	0,6241	0
28	0,6113	0
29	0,5956	0
30	0,7068	0
31	0,6119	0
32	0,5937	0
33	0,6738	0
34	0,6544	0
35	0,7156	0
36	0,6194	0
37	0,6212	0
38	0,6151	0
39	0,6599	0
40	0,6983	0
Total	26,1661	31,0434

 $\beta = 0,6$ 

Tipo	Apuestas totales	Ganancia
1	0,9383	0
2	0,6335	0
3	0,5593	0
4	0,8625	0
5	0,9036	0
6	0,8644	0
7	0,9158	0
8	0,9457	0
9	0,9167	23,3719
10	0,9631	0
11	0,7003	0
12	0,9397	0
13	0,4366	0
14	0,6087	0
15	0,9039	0
16	0,9563	0
17	0,6083	0
18	0,984	0
19	0,9896	0
20	0,8687	0
21	0,9869	0
22	0,8146	0
23	0,5776	0
24	0,9519	0
25	0,8912	0
26	0,7785	0
27	0,8898	0
28	0,4367	0
29	0,8396	0
30	0,5924	0
31	0,951	0
32	0,9548	0
33	0,9468	0
34	0,9596	0
35	0,6011	0
36	0,8731	0
37	0,9448	0
38	0,9508	0
39	0,9459	0
40	0,5694	0
Total	32,9555	23,3719

**Tabla 3.28.** *Leave-One-Out*: apuestas totales y ganancias para cada tipo, con  $\beta = 0,4$  y  $\beta = 0,6$

- $\beta = 0,4$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0	0,6034	0	0	0	
a2	0	0	0	0	0	0	0	0	0	0,7346						
a3	0	0	0	0	0	0	0	0	0,6868	0						
a4	0	0	0	0	0	0	0	0	0	0,6066	0	0	0			
a5	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6163		
a6	0	0	0	0	0	0	0	0	0	0	0	0,6026	0	0		
a7	0	0	0	0	0	0	0	0	0	0	0	0	0,656			
a8	0	0	0	0	0	0	0	0	0	0	0	0	0,6189	0		
a9	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6087		
a10	0	0	0	0	0	0	0	0	0	0,642						
a11	0	0	0	0	0	0	0	0	0	0,7845						
a12	0	0	0	0	0	0	0	0	0	0	0	0	0,6385	0		
a13	0	0	0	0	0	0	0	0	0	0	0	0	0,6035	0		
a14	0	0	0	0	0	0	0	0	0	0	0,7141					
a15	0	0	0	0	0	0	0	0,6105	0	0	0	0	0	0	0	
a16	0	0	0	0	0	0	0	0	0	0	0	0	0,6642			
a17	0	0	0	0	0	0	0	0	0	0	0,7198	0	0			
a18	0	0	0	0	0	0	0	0	0	0,6021						
a19	0	0	0	0	0	0	0	0	0	0,6242						
a20	0,584	0	0	0	0	0	0	0	0	0						
a21	0	0	0	0	0	0	0	0	0	0	0	0	0,6812	0		
a22	0	0	0	0	0	0	0	0	0,5916	0	0	0	0	0	0	0
a23	0	0	0	0	0	0	0	0	0,6973	0	0					
a24	0	0	0	0	0	0	0,5655	0	0	0	0	0	0	0	0	
a25	0	0,606	0	0	0	0	0	0	0	0	0	0	0	0	0	
a26	0	0	0	0	0	0	0	0	0	0,8321						
a27	0	0	0	0	0	0	0	0,6241	0	0	0	0	0	0	0	
a28	0	0	0	0	0	0	0	0	0	0,6113	0	0	0	0	0	
a29	0	0	0	0	0	0	0	0	0	0	0,5956	0				
a30	0	0	0	0	0	0	0	0,7068	0	0						
a31	0	0	0	0	0	0	0	0	0	0,6119	0	0	0	0		
a32	0	0	0	0	0	0	0	0	0,5937	0	0	0				
a33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6738	
a34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6544
a35	0	0	0	0	0	0	0	0	0,7156	0						
a36	0	0	0	0	0	0	0	0	0	0,6194						
a37	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6212		
a38	0	0	0	0	0	0	0	0	0	0,6151	0	0	0	0		
a39	0	0	0	0	0	0	0	0	0	0	0,6599	0	0	0		
a40	0	0	0	0	0	0	0	0	0,6983							

**Tabla 3.29.** *Leave-One-Out*: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,4$

- $\beta = 0,6$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0	0,4393	0,499	0	0	
a2	0	0	0	0	0	0	0	0	0	0,6335						
a3	0	0	0	0	0	0	0	0	0,5593	0						
a4	0	0	0	0	0	0	0	0	0	0,4489	0	0,4135	0			
a5	0	0	0	0	0	0	0	0	0	0	0	0,4434	0	0,4602		
a6	0	0	0	0	0	0	0	0	0	0	0	0,4381	0	0,4263		
a7	0	0	0	0	0	0	0	0,398	0	0	0	0	0,5179			
a8	0	0	0	0	0	0	0	0	0	0	0	0	0,4753	0,4704		
a9	0	0	0	0	0	0	0	0	0	0	0	0,4584	0	0,4583		
a10	0	0	0,4452	0	0	0	0	0	0	0,5179						
a11	0	0	0	0	0	0	0	0	0	0,7003						
a12	0	0	0	0	0	0	0	0	0	0	0	0	0,4983	0,4414		
a13	0	0	0	0	0	0	0	0	0	0	0	0	0,4366	0		
a14	0	0	0	0	0	0	0	0	0	0	0,6087					
a15	0	0	0	0	0	0	0	0,47	0	0	0	0	0	0,4339		
a16	0	0	0	0	0	0	0	0	0	0	0,4235	0	0,5328			
a17	0	0	0	0	0	0	0	0	0	0	0	0,6083	0	0		
a18	0	0	0	0	0	0,5352	0	0	0	0,4488						
a19	0	0	0,4965	0	0	0	0	0	0	0,4932						
a20	0,4134	0	0	0	0	0	0	0	0	0,4553						
a21	0	0	0	0	0	0	0	0	0	0	0	0	0,5598	0,4271		
a22	0	0	0	0	0	0	0	0	0,4312	0	0	0	0	0,3833	0	0
a23	0	0	0	0	0	0	0	0	0,5776	0	0					
a24	0	0	0	0	0	0	0,4007	0	0	0	0	0	0,5512	0		
a25	0	0,4511	0	0	0	0	0	0	0	0	0	0	0	0,4401		
a26	0	0	0	0	0	0	0	0	0	0,7785						
a27	0	0	0	0	0	0	0	0,4694	0	0	0	0,4204	0	0		
a28	0	0	0	0	0	0	0	0	0	0,4367	0	0	0	0		
a29	0	0	0	0,4053	0	0	0	0	0	0	0,4343	0				
a30	0	0	0	0	0	0	0	0,5924	0	0						
a31	0	0	0	0	0	0	0	0	0	0,4549	0	0	0	0,4961		
a32	0	0	0	0	0	0	0	0	0,4212	0	0,5336	0				
a33	0	0	0	0	0	0	0,3857	0	0	0	0	0	0	0	0,5611	
a34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,4441	0,5156
a35	0	0	0	0	0	0	0	0	0,6011	0						
a36	0	0	0	0	0	0	0,4134	0	0	0,4597						
a37	0	0	0	0	0	0	0	0	0	0	0,4609	0	0	0,4838		
a38	0	0	0	0	0	0	0	0	0	0,4544	0,4965	0	0	0		
a39	0	0	0	0	0	0	0	0	0	0	0,5304	0	0,4154			
a40	0	0	0	0	0	0	0	0	0,5694							

**Tabla 3.30.** *Leave-One-Out*: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,6$



$\beta = 0,7$ 

Tipo	Apuestas totales	Ganancia
1	0,7769	0
2	0,583	0
3	0,8824	0
4	0,6831	0
5	0,7339	0
6	0,6827	0
7	0,7451	0
8	0,7918	0
9	0,7509	19,2163
10	0,7939	0
11	0,985	0
12	0,7808	0
13	0,8733	0
14	0,9384	0
15	0,739	0
16	0,7993	0
17	0,9359	0
18	0,8125	0
19	0,8302	0
20	0,6848	0
21	0,8335	0
22	0,9717	0
23	0,9227	0
24	0,789	0
25	0,7222	0
26	0,7517	0
27	0,7163	0
28	0,8579	0
29	0,6573	0
30	0,884	0
31	0,7895	0
32	0,7889	0
33	0,8775	0
34	0,8005	0
35	0,9222	0
36	0,6911	0
37	0,7943	0
38	0,7865	0
39	0,7853	0
40	0,9576	0
Total	32,1026	19,2163

 $\beta = 0,8$ 

Tipo	Apuestas totales	Ganancia
1	0,8862	0
2	0,8985	0
3	0,7283	0
4	0,768	6,5019
5	0,8282	0
6	0,8259	0
7	0,8693	5,864
8	0,8721	0
9	0,8247	15,3004
10	0,9054	0
11	0,8381	0
12	0,915	0
13	0,9631	0
14	0,7971	0
15	0,8566	0
16	0,9781	0
17	0,7921	0
18	0,9654	0
19	0,9697	0
20	0,8774	0
21	0,9174	0
22	0,6915	0
23	0,7801	0
24	0,878	0
25	0,809	0
26	0,9644	0
27	0,8232	0
28	0,9314	9,1194
29	0,7494	0
30	0,7376	0
31	0,9206	0
32	0,9539	0
33	0,9546	0
34	0,8876	0
35	0,771	0
36	0,8245	0
37	0,9576	15,5275
38	0,9597	0
39	0,9159	0
40	0,8156	0
Total	34,6022	52,3132

**Tabla 3.31.** *Leave-One-Out*: apuestas totales y ganancias para cada tipo, con  $\beta = 0,7$  y  $\beta = 0,8$

- $\beta = 0,7$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0	0,354	0,4229	0	0	
a2	0	0	0	0	0	0	0	0	0	0,583						
a3	0	0	0	0	0	0	0	0	0,5069	0,3755						
a4	0	0	0	0	0	0	0	0	0,3626		0	0,3205	0			
a5	0	0	0	0	0	0	0	0	0	0	0	0,3564	0	0,3775		
a6	0	0	0	0	0	0	0	0	0	0	0	0,3497	0	0,333		
a7	0	0	0	0	0	0	0	0,3008	0	0	0	0	0,4443			
a8	0	0	0	0	0	0	0	0	0	0	0	0	0,3987	0,393		
a9	0	0	0	0	0	0	0	0	0	0	0	0,3741	0	0,3768		
a10	0	0	0,3545	0	0	0	0	0	0	0,4394						
a11	0	0	0	0	0	0	0,3136	0	0	0,6714						
a12	0	0	0	0	0	0	0	0	0	0	0	0	0,4236	0,3572		
a13	0	0	0	0	0	0	0	0	0	0	0	0	0,3583	0,515		
a14	0	0	0	0	0	0	0	0	0,3714	0	0,5669					
a15	0	0	0	0	0	0	0	0,3906	0	0	0	0	0	0,3484		
a16	0	0	0	0	0	0	0	0	0	0	0,3351	0	0,4642			
a17	0	0	0	0	0	0	0	0	0	0	0,5601	0,3758	0			
a18	0	0	0	0	0	0,4561	0	0	0	0,3564						
a19	0	0	0,4163	0	0	0	0	0	0	0,414						
a20	0,32	0	0	0	0	0	0	0	0	0,3648						
a21	0	0	0	0	0	0	0	0	0	0	0	0	0,4945	0,339		
a22	0	0	0	0	0	0	0	0	0,352	0	0	0,3235	0	0,2962	0	0
a23	0	0	0	0	0	0	0	0	0,5257	0,397	0					
a24	0	0	0	0	0	0	0,3878	0	0	0	0	0	0,4812	0		
a25	0	0,3675	0	0	0	0	0	0	0	0	0	0	0	0,3547		
a26	0	0	0	0	0	0	0	0	0	0,7517						
a27	0	0	0	0	0	0	0	0,3874	0	0	0	0,3289	0	0		
a28	0	0	0	0	0	0	0	0	0	0,3569	0	0	0	0,501		
a29	0	0	0	0,3095	0	0	0	0	0	0	0,3478	0				
a30	0	0	0	0	0	0	0	0,541	0	0,343						
a31	0	0	0	0	0	0	0	0	0	0,3715	0	0	0	0,418		
a32	0	0	0	0	0	0	0	0	0,3301	0	0,4588	0				
a33	0	0	0	0	0	0	0	0	0,3768	0	0	0	0	0	0,5007	
a34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,3574	0,4431
a35	0	0	0	0	0	0	0,3658	0	0,5564	0						
a36	0	0	0	0	0	0	0,3175	0	0	0,3736						
a37	0	0	0	0	0	0	0	0	0	0	0,3838	0	0	0,4105		
a38	0	0	0	0	0	0	0	0	0	0,3691	0,4174	0	0	0		
a39	0	0	0	0	0	0	0	0	0	0	0,4597	0	0,3256	0		
a40	0	0	0	0	0	0	0	0,4462	0,5114							

**Tabla 3.32.** *Leave-One-Out*: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,7$

- $\beta = 0,8$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a <sub>1</sub>	0	0	0	0	0	0	0	0	0	0	0	0,2714	0,3509	0	0,264	
a <sub>2</sub>	0	0	0	0	0	0	0	0	0,3607	0,5379						
a <sub>3</sub>	0	0	0	0	0	0	0	0	0,4405	0,2878						
a <sub>4</sub>	0	0	0	0	0	0	0	0	0	0,2827	0,2501	0,2353	0			
a <sub>5</sub>	0	0	0	0	0	0	0	0	0	0	0	0,2758	0,2536	0,2987		
a <sub>6</sub>	0	0	0	0	0	0	0	0	0	0,306	0	0,2677	0	0,2522		
a <sub>7</sub>	0	0	0	0	0	0,2792	0	0,2149	0	0	0	0	0,3752			
a <sub>8</sub>	0	0	0	0	0	0	0	0	0	0	0	0,2272	0,3257	0,3192		
a <sub>9</sub>	0	0	0	0	0	0	0	0	0	0	0	0,299	0,2257	0,3		
a <sub>10</sub>	0	0	0,2779	0	0	0	0	0	0,2526	0,3749						
a <sub>11</sub>	0	0	0	0	0	0	0,213	0	0	0,6251						
a <sub>12</sub>	0	0	0	0	0	0	0	0	0,2843	0	0	0	0,3533	0,2774		
a <sub>13</sub>	0	0	0	0	0	0	0,2299	0	0	0	0	0	0,2766	0,4566		
a <sub>14</sub>	0	0	0	0	0	0	0	0	0,2854	0	0,5117					
a <sub>15</sub>	0	0	0	0	0	0	0	0,3195	0,2658	0	0	0	0	0,2713		
a <sub>16</sub>	0	0	0	0	0	0	0	0	0	0	0,2534	0,326	0,3988			
a <sub>17</sub>	0	0	0	0	0	0	0	0	0	0	0,5026	0,2894	0			
a <sub>18</sub>	0	0	0	0	0	0,3932	0,2941	0	0	0,2781						
a <sub>19</sub>	0	0	0,351	0	0	0	0,272	0	0	0,3467						
a <sub>20</sub>	0,236	0	0	0	0	0	0,3483	0	0	0,2931						
a <sub>21</sub>	0	0	0	0	0	0	0	0	0,2291	0	0	0	0,4328	0,2555		
a <sub>22</sub>	0	0	0	0	0	0	0	0	0,2634	0	0	0,2286	0	0,1995	0	0
a <sub>23</sub>	0	0	0	0	0	0	0	0	0,4642	0,3159	0					
a <sub>24</sub>	0	0	0	0	0	0,2334	0,2224	0	0	0	0	0	0,4222	0		
a <sub>25</sub>	0	0,289	0	0	0	0	0	0	0	0	0	0,2458	0	0,2743		
a <sub>26</sub>	0	0	0	0	0	0	0,2355	0	0	0,7289						
a <sub>27</sub>	0	0	0	0	0	0	0	0,3097	0	0	0	0,2442	0,2693	0		
a <sub>28</sub>	0	0	0	0	0	0	0	0	0,2224	0,272	0	0	0	0,4371		
a <sub>29</sub>	0	0	0	0,2268	0,256	0	0	0	0	0	0,2666	0				
a <sub>30</sub>	0	0	0	0	0	0	0	0,4823	0	0,2553						
a <sub>31</sub>	0	0	0	0	0	0	0	0	0,2811	0,2924	0	0	0	0,3471		
a <sub>32</sub>	0	0	0	0	0	0	0	0	0,2436	0	0,3937	0,3166				
a <sub>33</sub>	0	0	0	0	0	0	0,2093	0	0,302	0	0	0	0	0	0,4432	
a <sub>34</sub>	0	0	0	0	0	0	0	0,2377	0	0	0	0	0	0	0,2768	0,373
a <sub>35</sub>	0	0	0	0	0	0	0,275	0	0,496	0						
a <sub>36</sub>	0	0	0	0	0	0	0,2321	0,2984	0	0,2939						
a <sub>37</sub>	0	0	0	0	0	0	0	0	0	0,3045	0,3113	0	0	0,3418		
a <sub>38</sub>	0	0	0	0	0	0	0	0	0	0,2888	0,345	0,3259	0	0		
a <sub>39</sub>	0	0	0	0	0	0	0	0	0	0	0,3948	0,2797	0,2415	0		
a <sub>40</sub>	0	0	0	0	0	0	0	0,3698	0,4458							

Tabla 3.33. *Leave-One-Out*: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,8$

$\beta = 0,9$ 

Tipo	Apuestas totales	Ganancia
1	0,9362	0
2	0,9181	0
3	0,7776	0
4	0,903	4,4103
5	0,9811	0
6	0,9442	0
7	0,8164	4,0797
8	0,9682	0
9	0,951	11,4369
10	0,8022	0
11	0,915	0
12	0,9394	0
13	0,9099	0
14	0,8911	0
15	0,8944	0
16	0,937	0
17	0,9136	0
18	0,9119	0
19	0,957	0
20	0,8358	0
21	0,847	0
22	0,9137	0
23	0,9331	0
24	0,7694	0
25	0,9114	0
26	0,8441	0
27	0,8212	0
28	0,9198	5,3296
29	0,7613	0
30	0,8083	0
31	0,8645	0
32	0,9627	0
33	0,9401	0
34	0,9495	0
35	0,8319	0
36	0,9596	0
37	0,7242	11,4312
38	0,9402	0
39	0,8737	0
40	0,9969	0
<b>Total</b>	<b>35,6757</b>	<b>36,6877</b>

 $\beta = 1$ 

Tipo	Apuestas totales	Ganancia
1	0,9281	0
2	0,8138	0
3	0,9418	0,3107
4	0,9297	2,2102
5	0,7956	0,2307
6	0,6788	0,615
7	0,6988	2,2307
8	0,7758	0,6761
9	0,6727	7,1689
10	0,7721	0,0112
11	0,8445	0,9408
12	0,858	0
13	0,905	0,0506
14	0,65	0
15	0,6115	0,113
16	0,942	0,9764
17	0,7495	0
18	0,9298	0,7405
19	0,943	0,5114
20	0,9541	0
21	0,9576	0,2367
22	0,4738	0
23	0,9605	0
24	0,7512	0
25	0,8772	0
26	0,9178	1,0491
27	0,8358	0,2048
28	0,8697	1,5014
29	0,8269	0,423
30	0,9309	0,0545
31	0,9831	0
32	0,7789	0
33	0,8241	0,5111
34	0,7564	0,1447
35	0,7481	0,0885
36	0,6921	0,797
37	0,9744	7,8183
38	0,911	0
39	0,9824	0,5568
40	0,908	0
<b>Total</b>	<b>32,5981</b>	<b>30,1721</b>

**Tabla 3.34.** *Leave-One-Out*: apuestas totales y ganancias para cada tipo, con  $\beta = 0,9$  y  $\beta = 1$

- $\beta = 0,9$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a1	0	0	0	0	0	0	0	0	0	0	0,2896	0,1882	0,2779	0	0,1805	
a2	0	0	0	0	0	0	0,1402	0	0,2895	0,4884						
a3	0	0	0	0,1877	0	0	0	0	0,3801	0,2098						
a4	0	0	0,196	0	0	0	0	0,1819	0	0,204	0,1696	0,1515	0			
a5	0	0	0	0	0	0	0,1838	0	0	0,2072	0	0,1967	0,1726	0,2208		
a6	0	0,1888	0	0	0,166	0	0	0	0	0,2338	0	0,1852	0	0,1704		
a7	0	0	0	0	0	0,1943	0	0,1222	0	0	0,1965	0	0,3034			
a8	0	0	0,1493	0	0	0	0	0	0	0	0,1768	0,1437	0,2529	0,2455		
a9	0	0	0	0	0,1832	0	0	0	0	0	0,1742	0,2259	0,1434	0,2243		
a10	0	0	0,1906	0	0	0	0,3119	0	0	0,2997						
a11	0	0	0	0	0	0	0,1264	0	0,2008	0,5878						
a12	0	0	0	0	0	0	0,1438	0	0,204	0	0,1143	0	0,2813	0,196		
a13	0	0	0	0	0	0	0,1402	0	0	0	0,1808	0	0,1934	0,3955		
a14	0	0	0	0	0,2167	0	0	0	0,2112	0	0,4632					
a15	0	0	0,147	0	0	0	0	0,246	0,1861	0,1235	0	0	0	0,1918		
a16	0	0	0	0	0	0	0	0	0,1896	0	0,1673	0,2488	0,3313			
a17	0	0	0	0	0	0	0	0,131	0	0	0,4504	0,2133	0,1189			
a18	0	0	0	0	0	0,32	0,2085	0,1925	0	0,1909						
a19	0	0	0,2781	0	0	0	0,1892	0,2163	0	0,2734						
a20	0,1461	0	0	0	0	0	0,2713	0,2086	0	0,2098						
a21	0	0	0	0	0	0	0	0	0,1393	0	0,1699	0	0,3687	0,1691		
a22	0	0	0	0	0	0	0	0	0,1856	0,2449	0	0,1485	0	0,1138	0	0,2209
a23	0	0	0,1404	0	0	0	0	0	0,4085	0,2434	0,1408					
a24	0	0	0	0	0	0,1431	0,1311	0	0	0	0	0,1407	0,3545	0		
a25	0	0,2107	0	0	0	0	0	0,216	0	0,1273	0	0,1632	0	0,1942		
a26	0	0	0	0	0	0	0,1436	0	0	0,7005						
a27	0	0	0	0	0	0	0	0,2303	0	0,2489	0	0,1568	0,1852	0		
a28	0	0	0	0	0	0	0	0	0,13	0,1857	0,2326	0	0	0,3715		
a29	0	0	0	0,1403	0,1736	0	0	0	0	0,1302	0,1838	0,1334				
a30	0	0	0	0	0	0,2086	0	0,4272	0	0,1725						
a31	0	0	0	0	0	0	0	0	0,1973	0,2106	0,1849	0	0	0,2717		
a32	0	0	0	0	0	0	0	0,2464	0,155	0	0,324	0,2373				
a33	0	0	0	0	0	0	0,1208	0	0,225	0	0	0	0	0,2104	0,3839	
a34	0	0	0	0	0	0	0	0,1534	0,1483	0	0,1495	0	0	0	0,1956	0,3027
a35	0	0	0	0	0	0	0,1961	0	0,4427	0,1931						
a36	0	0	0	0,1752	0	0,2004	0,1468	0,223	0	0,2142						
a37	0	0	0	0	0	0	0	0	0	0,2241	0,2329	0	0	0,2672		
a38	0	0	0	0	0	0	0	0	0	0,2056	0,2688	0,2473	0	0,2185		
a39	0	0	0	0	0	0	0	0	0	0	0,3264	0,1968	0,1539	0,1966		
a40	0	0	0,1763	0	0	0,1342	0	0,3016	0,3848							

Tabla 3.35. *Leave-One-Out*: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 0,9$

- $\beta = 1$

	Caballo1	Caballo2	Caballo3	Caballo4	Caballo5	Caballo6	Caballo7	Caballo8	Caballo9	Caballo10	Caballo11	Caballo12	Caballo13	Caballo14	Caballo15	Caballo16
a <sub>1</sub>	0,0103	0	0	0,0412	0,008	0,0025	0	0,0369	0,0894	0,0362	0,2171	0,1045	0,2041	0,0822	0,0958	
a <sub>2</sub>	0	0	0,0181	0,0056	0	0,0509	0,047	0,0405	0,2146	0,437						
a <sub>3</sub>	0	0,0565	0,0147	0,1148	0,0465	0,0643	0,1607	0,0241	0,323	0,1372						
a <sub>4</sub>	0	0	0,1147	0,1232	0,0384	0,021	0,0678	0,099	0,0292	0,1227	0,085	0,0645	0,1641			
a <sub>5</sub>	0	0	0,0355	0,0234	0	0,0179	0,0908	0,0469	0,134	0,1184	0	0,1093	0,0816	0,1378		
a <sub>6</sub>	0,0946	0,087	0	0	0,0616	0,0276	0,0504	0,0159	0	0,1391	0,022	0,0952	0,0116	0,0738		
a <sub>7</sub>	0	0	0	0,0502	0	0,1062	0,0765	0,027	0,029	0,0043	0,1068	0,0681	0,2307			
a <sub>8</sub>	0	0	0,0545	0,0425	0,0615	0,0158	0,084	0,0144	0	0,0229	0,085	0,0517	0,1757	0,1676		
a <sub>9</sub>	0,0061	0,0282	0	0,0285	0,0873	0,0288	0,0119	0,0034	0,0777	0	0,0773	0,1373	0,0456	0,1406		
a <sub>10</sub>	0	0,0014	0,1069	0,0462	0	0,0255	0,2414	0,0474	0,0752	0,2282						
a <sub>11</sub>	0	0,0132	0,0228	0,036	0	0	0,0305	0,0855	0,112	0,5446						
a <sub>12</sub>	0,0044	0	0	0	0,0139	0,0085	0,0539	0,087	0,1212	0,0677	0,0215	0,1592	0,2078	0,1129		
a <sub>13</sub>	0	0,0113	0,031	0,0115	0,0564	0	0,0517	0,0364	0,0314	0,0092	0,0969	0,1232	0,1107	0,3354		
a <sub>14</sub>	0	0	0	0	0,1214	0	0	0,0033	0,1202	0	0,4051					
a <sub>15</sub>	0,0164	0	0,0482	0	0,019	0	0,1036	0,1632	0,0941	0,0246	0	0,0019	0,0378	0,1029		
a <sub>16</sub>	0	0	0,0084	0,0236	0,0066	0,0498	0,0769	0,0856	0,107	0,0651	0,0821	0,1727	0,2642			
a <sub>17</sub>	0	0	0	0,0218	0,0198	0,002	0,1006	0,0333	0,0334	0	0,3925	0,1262	0,0199			
a <sub>18</sub>	0	0,0741	0,113	0,0484	0,0822	0,2504	0,1264	0,1087	0,0198	0,1067						
a <sub>19</sub>	0,0823	0,0639	0,2051	0	0,0439	0,0791	0,1063	0,1363	0,026	0,2						
a <sub>20</sub>	0,0603	0	0,0674	0,1053	0,0293	0,1691	0,2051	0,1368	0,0464	0,1344						
a <sub>21</sub>	0	0	0,0415	0,0814	0,0123	0,0315	0,0215	0,0653	0,0531	0,0286	0,0876	0,1412	0,3074	0,0861		
a <sub>22</sub>	0	0	0	0,0035	0,0417	0,0073	0	0	0,0934	0,1434	0,0057	0,049	0	0,0136	0,0011	0,1151
a <sub>23</sub>	0,018	0,0617	0,0538	0	0,0251	0,0965	0,0629	0,0706	0,3504	0,1673	0,0542					
a <sub>24</sub>	0	0,0019	0,0479	0,0343	0,0295	0,0558	0,0425	0,0147	0,1085	0,0114	0,0075	0,0532	0,2908	0,0531		
a <sub>25</sub>	0,0837	0,1298	0,033	0,0378	0	0	0,0512	0,1347	0,1253	0,0366	0,0108	0,0767	0,0462	0,1114		
a <sub>26</sub>	0	0	0,0099	0,0807	0,0062	0,0011	0,0581	0,0289	0,0575	0,6754						
a <sub>27</sub>	0	0	0,0299	0,0648	0,0273	0,0299	0,0173	0,1512	0	0,1727	0,0805	0,0699	0,1015	0,0908		
a <sub>28</sub>	0	0,0094	0,0124	0,0444	0	0,0094	0,0394	0,0174	0,0366	0,099	0,15	0,0286	0,118	0,3051		
a <sub>29</sub>	0,107	0	0,2248	0,0563	0,0943	0,046	0,0564	0,0427	0	0,0469	0,1021	0,0504				
a <sub>30</sub>	0	0,0404	0,0444	0,0078	0,08	0,1325	0,0426	0,373	0,1195	0,0907						
a <sub>31</sub>	0,053	0	0,0403	0,0334	0,0212	0,1103	0,0371	0,0108	0,1179	0,1313	0,1048	0,0307	0,0917	0,2005		
a <sub>32</sub>	0,0117	0	0	0,0147	0,0013	0,0046	0,033	0,1591	0,0636	0,0899	0,2487	0,1523				
a <sub>33</sub>	0	0	0,0341	0,0502	0,028	0,0105	0,0303	0,009	0,1455	0,0063	0,0184	0,0264	0,016	0,1268	0,3226	
a <sub>34</sub>	0	0,0148	0,0096	0,0233	0,0214	0	0	0,0569	0,0512	0,0893	0,0526	0,0342	0,0414	0,0235	0,1082	0,2299
a <sub>35</sub>	0,0134	0	0,0783	0	0	0,0148	0,1091	0,0427	0,3846	0,1053						
a <sub>36</sub>	0	0,0117	0	0,0764	0,0191	0,1033	0,0481	0,13	0,1771	0,1264						
a <sub>37</sub>	0	0	0,0198	0,0077	0,0028	0,0073	0,0324	0,1154	0,046	0,1533	0,1617	0,1996	0,0285	0,1999		
a <sub>38</sub>	0	0	0	0,028	0,0167	0,0538	0,0166	0,0177	0,0614	0,1223	0,1925	0,1685	0,097	0,1365		
a <sub>39</sub>	0,0196	0	0,0795	0,0571	0,0304	0,0721	0,0023	0,0837	0,0106	0,0611	0,261	0,1175	0,0694	0,118		
a <sub>40</sub>	0,131	0	0,0851	0,0937	0	0,0384	0,0132	0,2265	0,32							

**Tabla 3.36.** *Leave-One-Out*: apuestas que realizamos para cada caballo, en función del tipo, con  $\beta = 1$

Como vemos, la diferencia más grande entre lo que apostamos y lo que ganamos se da con  $\beta = 0,8$ . Con ese valor tenemos un beneficio de 17,71€. Por lo tanto, en este caso podríamos decir que el valor de  $\beta$  óptimo sería 0,8.

Además, al contrario que en el caso de validación cruzada, que con el valor que nos aportaba más beneficio sólo apostábamos por un caballo en cada carrera, con  $\beta = 0,8$ , estamos apostando por dos o tres caballos, dependiendo de la carrera.

Con  $\beta = 0,9$  apostamos por tres o cuatro caballos, en función de la carrera, y obtenemos un beneficio de 1€. Si lo comparamos con los 17€ de  $\beta = 0,8$  vemos que es un valor muy bajo. Aún así, tendríamos más posibilidades de acertar que en el caso de que  $\beta$  tome el valor de 0,8.

Sin embargo, en este caso no es tan arriesgado el utilizar el valor de 0,8, y además estamos teniendo una ganancia muy grande. Por lo tanto, en este caso podríamos concluir que el valor de  $\beta$  óptimo sería 0,8 independientemente del tipo de apostante que seamos.





## **CAPÍTULO 4**

# **Conclusiones y futuras líneas de trabajo**

En este capítulo vamos a tener dos cometidos:

- El primero es extraer las conclusiones a partir de los resultados mostrados en el capítulo anterior.
- Y el segundo es dar alguna pauta sobre cómo podría mejorarse, o ampliarse el Proyecto que estamos tratando.

## 4.1. Conclusiones

Como hemos podido ver en el capítulo anterior, dependiendo de los parámetros que empleemos para realizar la función SVM vamos a obtener beneficios más altos o más bajos e incluso pérdidas en algunas carreras.

Como estamos utilizando el Criterio de Kelly, vemos que cuando tenemos pérdidas, éstas son muy pequeñas, de forma que en las ganancias no se aprecia que repercuta de una forma muy significativa.

Hemos podido ver que, en el caso de que utilizáramos validación cruzada, existirían dos valores de  $\beta$  que serían óptimos dependiendo del tipo de apostante que seamos. Con  $\beta = 0$ , que significa que sólo vamos a tener en cuenta las probabilidades calculadas por *Formstar*, es decir, aquí la SVM no cumple ningún papel, obtendríamos más beneficios. En total tendríamos una ganancia de 49,8€, sin embargo, apostamos únicamente por un caballo en una carrera, lo que es demasiado arriesgado. Por lo tanto, bajo nuestro punto de vista sería mejor utilizar el valor de  $\beta = 0,9$ , ya que también nos reporta beneficios, aunque sean significativamente inferiores, y además, al apostar por un mayor número de caballos, tenemos más posibilidades de acertar el ganador.

Con  $\beta = 0,8$  y usando validación *Leave-One-Out* obtenemos una ganancia de 57,71€, que es el mayor beneficio que hemos obtenido de todas las pruebas realizadas. Además, en este caso estamos apostando por un número más alto de caballos por carrera, por lo que podríamos concluir que no sería tan arriesgado utilizar este valor.

Además, señalamos que entre los dos métodos utilizados, lo óptimo es utilizar *Leave-One-Out*, ya que con este método y  $\beta = 0,8$  maximizamos el beneficio, lo cuál es nuestro cometido. Esto es así debido a que con este método estamos utilizando más muestras de entrenamiento y, como hemos dicho, cuantas más carreras utilicemos para entrenar a la máquina, mejores resultados lograremos.

Indudablemente, el beneficio depende del *bankroll* del que dispongamos. Apostando 1€ por carrera obtenemos un beneficio total de 57,71€. Si en lugar de 1€, usáramos 10€ por carrera, la ganancia sería de 577,1€.

## 4.2. Futuras líneas de trabajo

Como mejoras para un Proyecto futuro basado en este, podemos proponer:

- Realizar los experimentos con una base de datos mayor. Si podemos entrenar nuestro programa con un mayor número de carreras, se ajustará más a los resultados de forma que con las nuevas carreras podamos obtener mayores beneficios. Además, es necesario ya que los resultados de LOO y CV no deberían ser muy diferentes, cosa que si lo son en nuestros resultados, por lo tanto necesitamos más datos para poder decir que el modelo funciona.
- También podríamos tener en cuenta otro tipo de información, como la posición de la clasificación en la que se encuentra el jinete o el entrenador.
- En nuestro caso hemos estudiado las carreras que tienen en común que se corren en un hipódromo de 1400 metros. También podríamos haber hecho la distinción en función del tipo de terreno en el que se realiza la carrera. E incluso, si tenemos una base de datos muy grande, podríamos combinar la longitud del hipódromo con el tipo de terreno.
- En este caso hemos realizado el análisis con SVM, otra opción sería emplear Procesos Gaussianos [70].
- Hemos manejado el tipo *Kernel* RBF, sin embargo hay más posibilidades como utilizar un *Kernel* lineal, cuadrático o polinómico.

Además, para evitar que se repita, vamos a indicar una línea de trabajo que intentamos seguir, y que no era correcta. Para disponer de un mayor número de carreras de entrenamiento, decidimos emplear todas las carreras que estuvieran entre los 1200 metros y los 1600 metros. Los resultados que nos dieron no fueron satisfactorios, ya que no obteníamos ganancias. Con esto nos hemos dado cuenta de que no es posible tratar de igual forma los caballos que corren en hipódromos con distintas longitudes.



# **CAPÍTULO 5**

## **Presupuesto**

En este capítulo vamos a abordar el presupuesto de este Proyecto Fin de Carrera. Para ello vamos a ir mostrando las distintas fases que ha llevado a que se pudiera hacer realidad, además del coste de material y personal que lleva asociado la puesta en marcha de este sistema.

Por último realizaremos un resumen del mismo, utilizando la plantilla proporcionada por la Universidad Carlos III.

## 5.1. Introducción

Las fases del trabajo las podemos dividir en:

- Fase 1. Análisis de la información y estudio de las distintas opciones disponibles.
- Fase 2. Creación de la base de datos.
- Fase 3. Desarrollo del modelo.
- Fase 4. Obtención de resultados.
- Fase 4. Documentación.

El tiempo total ha sido de 13 meses, con una distribución horaria variable entre las distintas fases seguidas para la realización. Se ha supuesto un esfuerzo continuo, aunque la línea de carga ha sido irregular a lo largo de todo el proceso. Desde el periodo estival hasta finalizar el Proyecto el tiempo empleado ha sido mayor que en periodos lectivos. Hemos seguido la siguiente distribución de horas entre fases:

- Fase 1. 90 días \* 2 horas/día = 180 horas
- Fase 2. 140 días \* 2 horas/día = 280 horas
- Fase 3. 50 días \* 3 horas/día = 150 horas
- Fase 4. 45 días \* 3 horas/día = 135 horas
- Fase 5. 65 días \* 3 horas/día = 195 horas

A continuación vamos a ir haciendo un desglose de los distintos costes asociados al Proyecto, para facilitar la lectura posterior de la plantilla proporcionada por la Universidad Carlos III de Madrid.

## 5.2. Coste del personal

El personal va a ser distinto en función de las fases que hemos llevado durante el Proyecto. Excepto en la Fase 2 y en la Fase 3, que en ambos casos ha hecho falta de un programador. Por lo tanto, para el cálculo del coste asociado al programador, vamos a sumar las horas de las dos fases mencionadas antes.

Puesto	Nº horas	Coste hora	Total (€)
Analista	180	30 €	5.400 €
Programador	430	25 €	10.750 €
Técnico de pruebas	135	22 €	2.970 €
Responsable documentación	195	15 €	2.925 €
			<b>22.045 €</b>

Tabla 5.1. Coste del personal

### 5.3. Coste de software

En este apartado incluiremos el coste de las aplicaciones y herramientas empleadas durante la realización del Proyecto, tanto para la programación como para la documentación del mismo.

Descripción	Coste (€)
Microsoft Office 2007 Professional	290 €
Licencia Matlab 2008	300 €
<b>590 €</b>	

Tabla 5.2. Coste de software

### 5.4. Coste de hardware

En este caso incluiremos todo el hardware que ha sido necesario para poder llevar a cabo el desarrollo del Proyecto. Todos los equipos utilizados eran propiedad del autor.

Descripción	Coste (€)	% Uso dedicado	Tiempo (meses)	Depreciación	Coste total (€)
PC HP	800 €	100%	13	60	173,33 €
Ratón y teclado HP	40 €	100%	13	60	8,67 €
Monitor TFT Samsung	150 €	100%	13	60	32,50 €
Pendrive Verbatim 4 Gb	12 €	100%	13	60	2,60 €
Impresora HP Deskjet 3050	89 €	100%	1	60	1,48 €
					<b>218,58 €</b>

Tabla 5.3. Coste de hardware

## 5.5. Coste de material fungible

Estos gastos incluyen el consumo de tinta de impresora realizado para imprimir documentación en la Fase 1 y en la Fase 5. Además también incluimos el material de oficina como son bolígrafos o cuadernos.

Descripción	Coste (€)
Cartuchos de tinta	100 €
Material de oficina	15 €
	<b>115 €</b>

**Tabla 5.4.** Coste de material fungible

## 5.6. Resumen de costes

Hacemos un resumen de los gastos calculados anteriormente, incluyendo los costes indirectos del 20% como se indica en la plantilla del presupuesto.

Descripción	Costes totales (€)
Personal	22.045 €
Software	590 €
Amortización del Hardware	219 €
Material fungible	115 €
Costes indirectos (20%)	4.594 €
	<b>27.562 €</b>

**Tabla 5.5.** Resumen de costes

A parte de este resultado final, hemos incluido una segunda tabla añadiendo un coste del 18% en concepto de impuestos sobre el valor añadido (IVA).

Descripción	Costes totales (€)
Total sin IVA	27.562 €
Total con IVA (18%)	4.961 €
	<b>32.523 €</b>

**Tabla 5.6.** Presupuesto total sin y con IVA

El presupuesto total de este Proyecto asciende a la cantidad de 32.523€.

La ingeniera proyectista,

Fdo. Miriam Trujillo Sevillano




 <b>UNIVERSIDAD CARLOS III DE MADRID</b> <b>Escuela Politécnica Superior</b>							
<b>PRESUPUESTO DE PROYECTO</b>							
<b>1.- Autor:</b>							
Miriam Trujillo Sevillano							
<b>2.- Departamento:</b>							
Teoría de la señal							
<b>3.- Descripción del Proyecto:</b>							
- Título	Predicción de resultados de eventos deportivos: Carreras de caballos						
- Duración (meses)	13						
Tasa de costes indirectos:	20%						
<b>4.- Presupuesto total del Proyecto (valores en Euros):</b>							
32.523,00 Euros							
<b>5.- Desglose presupuestario (costes directos)</b>							
<b>PERSONAL</b>							
Apellidos y nombre	N.I.F. (no rellenar - solo a título informativo)	Categoría	Dedicación (meses) <sup>a)</sup> (hombres)	Coste hombre mes	Coste (Euro)	Firma de conformidad	
Miriam Trujillo Sevillano		Analista	1,37	3.941,60	5.400,00		
Miriam Trujillo Sevillano		Programador	3,28	3.277,44	10.750,00		
Miriam Trujillo Sevillano		Técnico de pruebas	1,03	2.883,49	2.970,00		
Miriam Trujillo Sevillano		Documentación	1,49	1.963,09	2.925,00		
					0,00		
<b>Hombres mes 7,17</b>				<b>Total</b>	22.045,00		
<sup>a)</sup> 1 Hombre mes = 131,25 horas. Máximo anual de dedicación de 12 hombres mes (1575 horas) Máximo anual para PDI de la Universidad Carlos III de Madrid de 8,8 hombres mes (1.155 horas)							
<b>EQUIPOS</b>							
Descripción	Coste (Euro)	% Uso dedicado proyecto	Dedicación (meses)	Periodo de depreciación	Coste imputable <sup>d)</sup>		
PC HP	800 €	100	13	60	173,33		
Ratón y teclado HP	40 €	100	13	60	8,67		
Monitor TFT Samsung	150 €	100	13	60	32,50		
Pendrive Verbatim 4 Gb	12 €	100	13	60	2,60		
Impresora HP Deskjet 3050	89 €	100	1	60	1,48		
				<b>Total</b>	218,58		
<sup>d)</sup> Fórmula de cálculo de la Amortización: $\frac{A}{B} \times C \times D$ <p> <b>A</b> = nº de meses desde la fecha de facturación en que el equipo es utilizado  <b>B</b> = periodo de depreciación (60 meses)  <b>C</b> = coste del equipo (sin IVA)  <b>D</b> = % del uso que se dedica al proyecto (habitualmente 100%) </p>							
<b>SUBCONTRATACIÓN DE TAREAS</b>							
Descripción	Empresa	Coste imputable					
<b>Total</b>		0,00					
<b>OTROS COSTES DIRECTOS DEL PROYECTO<sup>e)</sup></b>							
Descripción	Empresa	Costes imputable					
Coste de Software		590,00					
Material fungible		115,00					
<b>Total</b>		705,00					
<sup>e)</sup> Este capítulo de gastos incluye todos los gastos no contemplados en los conceptos anteriores, por ejemplo: fungible, viajes y dietas, otros,...							
<b>6.- Resumen de costes</b>							
Presupuesto Costes Totales	<b>Presupuesto Costes Totales</b>						
Personal	22.045						
Amortización	219						
Subcontratación de tareas	0						
Costes de funcionamiento	705						
Costes indirectos	4.594						
<b>Total</b>	<b>27.562</b>						

Tabla 5.7. Plantilla del presupuesto



# Referencias

## Páginas web

- [1] <http://formstar.net/racing/index.php>
- [2] <http://www.apostar.ws/historia.php>
- [3] <http://www.retrofootball.com/blog/204/>
- [4] <http://concursodeapuestas.com/historia-de-las-apuestas-deportivas/>
- [5] <http://www.apuestascaballos.net/glosario-carreras-caballos>
- [6] <http://www.rinconapuestas.com/diccionario/>
- [7] <http://www.apuestascaballos.net/apuestas-caballos/tipos-carrera-categoria-maiden-handicap-stakes-listed-claimers-sellers.html>
- [8] <http://www.apuestacaballos.com/tipos-de-carreras>
- [9] <http://tienespoker.com/apuestas-hipicas-online.php>
- [10] <http://www.elgalope.com/hipodromo.html>

- [11] <http://www.webapuestas.com/guia-tutorial-apuestas/apuestas-deportivas/caballos/guia-apuestas-caballos.html>
- [12] <http://www.apuestascaballos.net/tipos-de-apuestas/apuestas-al-tote-1-totewin-y-tote-place.html>
- [13] <http://www.webapuestas.com/guia-tutorial-apuestas/apuestas-deportivas/caballos/consejos-apuestas-caballos.html>
- [14] [http://elrecreativo.com/index.php?option=com\\_content&view=article&catid=110%3Anoticias-del-recreativo&id=10543%3Apais-vasco-el-perfil-del-apostante-varon-de-entre-25-y-45-anos-aficionado-al-deporte&Itemid=849](http://elrecreativo.com/index.php?option=com_content&view=article&catid=110%3Anoticias-del-recreativo&id=10543%3Apais-vasco-el-perfil-del-apostante-varon-de-entre-25-y-45-anos-aficionado-al-deporte&Itemid=849)
- [15] <http://www.baquia.com/posts/el-futbol-y-el-tenis-los-preferidos-de-los-apostantes-espanoles>
- [16] <http://www.lookandbet.com/2009/03/perfil-del-usuario-de-apuestas-online.html>
- [17] <http://armasdejuego.com/apuestas-deportivas/estrategias-trucos-y-consejos/el-perfil-de-un-apostante.html>
- [18] <http://www.soloapuestasdeportivas.es/%C2%BFcomo-funcionan-las-casas-de-apuestas/7>
- [19] <http://capital2max.com/archives/508>
- [20] [http://es.wikipedia.org/wiki/Corredor\\_de\\_apuestas](http://es.wikipedia.org/wiki/Corredor_de_apuestas)
- [21] <http://apuestasdecaballos.com.es/apuesta-con-zeturf-primera-casa-de-apuestas-hipicas-on-line/>
- [22] <http://www.zeturf.com/es/>
- [23] <http://apostarcarrerascaballosfrancia.blog.com.es/2011/12/04/como-apostar-en-las-carreras-francesas-12254295/>
- [24] <http://www.webapuestas.com/casas-apuestas/betfair/index.html>

- [25] <http://www.betfair.com/es/>
- [70] [http://es.wikipedia.org/wiki/Validaci%C3%B3n\\_cruzada](http://es.wikipedia.org/wiki/Validaci%C3%B3n_cruzada)
- [71] [http://es.wikipedia.org/wiki/M%C3%A1quinas\\_de\\_vectores\\_de\\_soporte](http://es.wikipedia.org/wiki/M%C3%A1quinas_de_vectores_de_soporte)
- [72] <http://www.mathworks.es/help/toolbox/bioinfo/ref/svmtrain.html>
- [73] <http://www.webapuestas.com/guia-tutorial-apuestas/gestion-bankroll/criterio-kelly.html>

## Artículos

- [26] Searching for positive returns at the track: a multinomial logit model for handicapping horse races. Bolton R.N., Chapman, R.G., 1986.
- [27] Financial Theory and Corporate Policy. Copeland, T.E., Weston, J.F., 1979,
- [28] Efficient Capital Markets: A Review of Theory and Empirical Work. Fama, E.F., 1970
- [29] Securities Market Efficiency in an Arrow-Debreu Market. Rubinstein, M., 1975.
- [30] An empirical cross-validation of alternative classification strategies applied to harness racing data for win bets. Ludlow, L.H.
- [31] Estimating the probabilities of the outcomes of a horse race (alternatives to the Harville formulas). Stern, H.S. [46] Adapting support vector machine methods for horserace odds prediction. Edelman, D., 2007
- [32] Racetrack betting and informed behaviour. Peter Asch, Burton G. Malkiel and Richard E. Quandt, 1982.
- [33] Identifying winners of competitive events: A SVM-based classification model for horserace prediction. Stefan Lessman, Ming-Chien Sung, Johnnie E.V. Johnson, 2009.
- [34] Transactions costs, market inefficiencies and entries in a racetrack betting model. Hausch, D.B., Ziemba, W.T., 1985

- [35] Exploring decision makers' use of price information in a speculative market. Johnson, J.E.V., Jones, O., Tang, L., 2006
- [36] Insider trading, herding behaviour and market plungers in the British horse-race betting market. Law, D., Peel, D.A., 2002
- [37] Why are gambling markets organised so differently from financial markets? Levitt, S.D., 2004
- [38] The economics of wagering markets. Sauer, R.D., 1998
- [39] Inside information in a betting market. Schnytzer, A., Shilony, Y., 1995
- [40] Information efficiency in betting markets. Vaughan Williams, L., 1999
- [41] Computer based horse race handicapping and wagering systems: A report. Benter, W., 1994
- [42] Searching for semi-strong form inefficiency in the UK racetrack betting market. Sung, M.C., Johnson, J.E.V., Bruce, A.C., 2005
- [43] Conditional logit analysis of qualitative choice behaviour. McFadden, D., 1973
- [44] Still searching for positive returns at the track: Empirical results from 2000 Hong Kong races. Chapman, R.G., 1994
- [45] Multinomial Probit Models for Competitive Horse Racing. Gu, M.G., Huang, C., Benter, W., 2003
- [46] Adapting support vector machine methods for horserace odds prediction. Edelman, D., 2007
- [47] Advances in the mathematical modelling of horse race outcomes. Benter, W., 2003
- [48] Comparing the effectiveness of one and two step conditional logit models for predicting outcomes in a speculative market. Sung, M., Johnson, J.E.V., 2007
- [49] The Nature of Statistical Learning Theory. Vapnik, V.N., 1995

- [50] An introduction to Support Vector Machines and other Kernel-based Learning Methods. Cristianini, N., Shawe-Taylor, J., 2000
- [51] A tutorial on support vector machines for pattern recognition. Burges, C.J.C., 1998
- [52] Benchmarking least squares support vector machine classifiers. Van Gestel, T., Suykens, J.A.K., Baesens, B., Viane, S., Vanthienen, J., Dedene, G., De Moor, B., Vandewalle, J., 2004
- [53] Predictions in marketing using the support vector machine. Cui, D., Curry, D., 2005
- [54] Least square support vector machine classifiers. Suykens, J.A.K., Vanderwalle, J., 1999
- [55] Subjective information and market efficiency in a betting market. Figlewski, S., 1979
- [56] Investigating the roots of the favourite-longshot bias: An analysis of supply and demand side agents in parallel betting markets. Bruce, A.C., Johnson, J.E.V., 2000
- [57] A tutorial on support vector regression. Smola, A.J., Schölkopf, B., 2004
- [58] Neural Networks for Pattern Recognition. Bishop, C.M., 1995
- [59] Cross-validation: choice and assessment of statistical predictions. Stone, M., 1974
- [60] A Practical Guide to Support Vector Classification. Hsu, C.W., Chang, C.C., Lin, C.J., 2003
- [61] A New Interpretation of Information Rate. Kelly, J.L., 1956.
- [62] An introduction to ROC analysis. Fawcett, T., 2006
- [63] Exploring rank ordered choice set data within the stochastic utility model. Chapman, R.G., Staelin, R., 1982
- [64] Transferability of disaggregated mode choice models. Watson, P.L., Westin, R.B., 1975

- [65] A comparison of methods for multi-class support vector machines. Hsu, C.W., Lin, C.J., 2002
- [66] Support vector learning for ordinal regression. Herbrich, R., Graepel, T., Obermayer, K., 1999
- [67] A fast dual algorithm for kernel logistic regression. Keerthi, S.S., Duan, K.B., Shevade, S.K., Poo, A.N., 2005
- [68] A support vector method for multivariate performance measures. Joachims, T., 2005
- [69] Comprehensible credit scoring models using rule extraction from support vector machines. Martens, D., Baesens, B., Van Gestel, T., Vanthienen, J., 2007
- [74] C. K. I. Williams and D. Barber. Bayesian classification with Gaussian processes, 1998.